GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: September 21, 2001, 00:35:06 ; Search time 3479.37 Seconds
(without alignments)
97.802 Million cell updates/sec

Title: US-09-138-735-9
Perfect score: 22
Sequence: 1 cagccgacggtagctgcgtcct 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*
44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vi1:*
59: gb_vi2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	22	100.0	22	9	A48918	A48918 Sequence 9
	2	22	100.0	22	9	AR047927	AR047927 Sequence
	3	22	100.0	3402	9	A48910	A48910 Sequence 1
	4	22	100.0	3402	9	AR047920	AR047920 Sequence
	5	22	100.0	3402	96	TCU24190	U24190 Trypanosoma
c	6	17.8	80.9	158566	66	AC021683	AC021683 Homo sapi
	7	17.2	78.2	29485	4	AC005439	AC005439 Drosophil
c	8	17.2	78.2	33077	64	AC017302	AC017302 Drosophil

	9	17.2	78.2	36676	93	HSJ858B16 AL096768 Human DNA
c	10	17.2	78.2	97308	63	AC013925 AC013925 Drosophil
c	11	17.2	78.2	151369	13	AP001080 AP001080 Oryza sat
c	12	17.2	78.2	161046	58	AF081810 AF081810 Lymantria
c	13	17.2	78.2	168655	13	AP000969 AP000969 Oryza sat
	14	17.2	78.2	175682	4	AC007417 AC007417 Drosophil
	15	17.2	78.2	261846	5	AE003830 AE003830 Drosophil
	16	16.8	76.4	1188	14	GB58S X93990 G.biloba DN
c	17	16.8	76.4	66915	63	AC013073 AC013073 Drosophil
c	18	16.8	76.4	117486	83	AP003037 AP003037 Homo sapi
	19	16.8	76.4	127652	82	AP001526 AP001526 Homo sapi
c	20	16.8	76.4	152936	60	AC004229 AC004229 Homo sapi
c	21	16.8	76.4	222472	69	AC026091 AC026091 Homo sapi
c	22	16.4	74.5	11475	1	AE005794 AE005794 Caulobact
c	23	16.4	74.5	107600	13	ATF16J13 AL049638 Arabidops
	24	16.4	74.5	110000	84	LMFLCHR32_15 Continuation (16 o
c	25	16.4	74.5	190026	13	ATCHRIV33 AL161533 Arabidops
	26	16.2	73.6	2426	1	AF047430 AF047430 Sinorhizo
	27	16.2	73.6	2429	89	AK026936 AK026936 Homo sapi
	28	16.2	73.6	4051	88	AF006005 AF006005 Homo sapi
	29	16.2	73.6	5638	91	D83784 D83784 Human mRNA
	30	16.2	73.6	10160	1	AE002024 AE002024 Deinococc
c	31	16.2	73.6	10282	1	AE005100 AE005100 Halobacte
	32	16.2	73.6	10413	1	AE005018 AE005018 Halobacte
c	33	16.2	73.6	10475	10	I08101 I08101 Sequence 4
c	34	16.2	73.6	10475	10	I08711 I08711 Sequence 9
c	35	16.2	73.6	10476	97	HUMSPBAA M24461 Human pulmo
	36	16.2	73.6	24008	64	AC017420 AC017420 Drosophil
c	37	16.2	73.6	28272	6	CET07A5 Z48055 Caenorhabdi
c	38	16.2	73.6	38397	6	DMC8D8 AL022018 Drosophil
c	39	16.2	73.6	54327	78	AC090435 AC090435 Chlamydom
c	40	16.2	73.6	58198	78	AC091323 AC091323 Mus muscu
	41	16.2	73.6	60135	78	AC091200 AC091200 Drosophil
c	42	16.2	73.6	62148	78	AF192524 AF192524 Homo sapi
c	43	16.2	73.6	88050	78	AC090436 AC090436 Chlamydom
	44	16.2	73.6	104171	65	AC019499 AC019499 Drosophil
	45	16.2	73.6	110098	90	AL358976 AL358976 Human DNA

ALIGNMENTS

```
RESULT   1
A48918
LOCUS       A48918          22 bp    DNA             PAT       07-MAR-1997
DEFINITION  Sequence 9 from Patent WO9605312.
ACCESSION   A48918
VERSION     A48918.1  GI:2302577
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 22)
  AUTHORS   Paranhos-Baccala,G., Lesenechal,M. and Jolivet,M.
  TITLE     NOVEL TRYPANOSOMA CRUZI ANTIGEN, AND GENE CODING THEREFOR; THEIR
            APPLICATION TO THE DETECTION OF CHAGAS' DISEASE
  JOURNAL   Patent: WO 9605312-A 9 22-FEB-1996;
            BIO MERIEUX (FR)
COMMENT     Other publication CA 2173957 960222
            Other publication AU 3169195 960307
            Other publication FR 2723589 960216.
FEATURES             Location/Qualifiers
     source          1. .22
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT        3 a      8 c      7 g      4 t
ORIGIN


  Query Match           100.0%;  Score 22;  DB 9;  Length 22;
  Best Local Similarity  100.0%;  Pred. No. 15;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcgtcct 22
          ||||||||||||||||||||||
Db      1 CAGCCGACGGTAGCTGCGTCCT 22


RESULT   2
AR047927
LOCUS       AR047927       22 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 9 from patent US 5820864.
ACCESSION   AR047927
VERSION     AR047927.1  GI:5970270
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 22)
  AUTHORS   Paranhos-Baccala,G., Lesenechal,M. and Jolivet,M.
  TITLE     Trypanosoma cruzi antigen, gene encoding therefor and methods of
            detecting and treating chagas disease
  JOURNAL   Patent: US 5820864-A 9 13-OCT-1998;
FEATURES             Location/Qualifiers
     source          1. .22
                     /organism="unknown"
BASE COUNT        3 a      8 c      7 g      4 t
ORIGIN


  Query Match           100.0%;  Score 22;  DB 9;  Length 22;
  Best Local Similarity  100.0%;  Pred. No. 15;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcgtcct 22
          ||||||||||||||||||||||
Db      1 CAGCCGACGGTAGCTGCGTCCT 22


RESULT   3
A48910
LOCUS       A48910       3402 bp    DNA             PAT       07-MAR-1997
DEFINITION  Sequence 1 from Patent WO9605312.
ACCESSION   A48910
VERSION     A48910.1  GI:2302570
KEYWORDS    .
SOURCE      Trypanosoma cruzi.
  ORGANISM  Trypanosoma cruzi
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma; Schizotrypanum.
REFERENCE   1  (bases 1 to 3402)
  AUTHORS   Paranhos-Baccala,G., Lesenechal,M. and Jolivet,M.
  TITLE     NOVEL TRYPANOSOMA CRUZI ANTIGEN, AND GENE CODING THEREFOR; THEIR
            APPLICATION TO THE DETECTION OF CHAGAS' DISEASE
  JOURNAL   Patent: WO 9605312-A 1 22-FEB-1996;
            BIO MERIEUX (FR)
COMMENT     Other publication CA 2173957 960222
            Other publication AU 3169195 960307
            Other publication FR 2723589 960216.
FEATURES             Location/Qualifiers
     source          1. .3402
                     /organism="Trypanosoma cruzi"
                     /strain="G"
                     /db_xref="taxon:5693"
                     /dev_stage="EPIMASTIGOTE"
BASE COUNT      889 a    818 c    958 g    737 t
ORIGIN


  Query Match           100.0%;  Score 22;  DB 9;  Length 3402;
  Best Local Similarity  100.0%;  Pred. No. 3.5;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcgtcct 22
          ||||||||||||||||||||||
Db   1266 CAGCCGACGGTAGCTGCGTCCT 1287
```

```
RESULT   4
AR047920
LOCUS       AR047920     3402 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 1 from patent US 5820864.
ACCESSION   AR047920
VERSION     AR047920.1  GI:5970263
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 3402)
  AUTHORS   Paranhos-Baccala,G., Lesenechal,M. and Jolivet,M.
  TITLE     Trypanosoma cruzi antigen, gene encoding therefor and methods of
            detecting and treating chagas disease
  JOURNAL   Patent: US 5820864-A 1 13-OCT-1998;
FEATURES             Location/Qualifiers
     source          1. .3402
                     /organism="unknown"
BASE COUNT      888 a    821 c    956 g    737 t
ORIGIN


  Query Match           100.0%;  Score 22;  DB 9;  Length 3402;
  Best Local Similarity  100.0%;  Pred. No. 3.5;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcgtcct 22
          ||||||||||||||||||||||
Db   1266 CAGCCGACGGTAGCTGCGTCCT 1287


RESULT   5
TCU24190
LOCUS       TCU24190     3402 bp    mRNA            INV       04-AUG-1997
DEFINITION  Trypanosoma cruzi Tc40 antigen (Tc40) mRNA, complete cds.
ACCESSION   U24190
VERSION     U24190.1  GI:790645
KEYWORDS    .
SOURCE      Trypanosoma cruzi.
  ORGANISM  Trypanosoma cruzi
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma; Schizotrypanum.
REFERENCE   1  (bases 1 to 3402)
  AUTHORS   Lesenechal,M., Duret,L., Cano,M.I., Mortara,R.A., Jolivet,M.,
            Camargo,M.E., da Silveira,J.F. and Paranhos-Baccala,G.
  TITLE     Cloning and characterization of a gene encoding a novel
            immunodominant antigen of Trypanosoma cruzi
  JOURNAL   Mol. Biochem. Parasitol. 87 (2), 193-204 (1997)
  MEDLINE   97391123
REFERENCE   2  (bases 1 to 3402)
  AUTHORS   Lesenechal,M., Franco Da Silveira,J., Mortara,R.A., Duret,L.,
            Camargo,M.E., Jolivet,M. and Paranhos-Baccala,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-APR-1995) Mylene Lesenechal, BioMerieux, 22 rue Saint
            Jean de Dieu, Lyon 69007, France
FEATURES             Location/Qualifiers
     source          1. .3402
                     /organism="Trypanosoma cruzi"
                     /strain="G"
                     /db_xref="taxon:5693"
                     /dev_stage="epimastigote"
     misc_RNA .      1. .35
                     /gene="Tc40"
                     /note="spliced leader"
     gene            1. .3402
                     /gene="Tc40"
     CDS             266. .3013
                     /gene="Tc40"
                     /codon_start=1
                     /product="Tc40 antigen"
                     /protein_id="AAC47657.1"
                     /db_xref="GI:790646"
                     /translation="MTVTVDLFNHAKPSNNEGRVWSVDAATFNEVPEAQRVLADSQFY
                     LAYTMKRRHVLRVVKRSNLLKGTVRAHSKPIHAVKFVNYRSNVAASAGKGEFFVWVVT
                     DETEASNGKPDLAARLTVKVYFKLQDPVTIPCFSFFINAESQRPDLLVLYETQAAILD
                     SSSLIERFDVESLEATLQRNCTTLRTLTQPVSENSLCSVGSGGWFTFTTEPTMVAACT
                     LRNRSTPSWACCEGEPVKALHLLDATVEENVSVLVAASTKGVYQWLLTGVAEPNLLRK
                     FVIDGSIVAMESSRETFAVFDDRKQLALVNMHSPHNFTCTHYMMPCQVQRNGFCFNRT
                     ADGSCVLADMSNRLTIFHLRSSRREEQQPGQKTSVVATAKPGCVSSGTDAASSSHTNT
                     TSAAAASPASPPVSAPAKAAAPPAAARSAEPHVGSKIIANLVNQLGINVTQRSVVSTG
                     APATTRSTAVTSTTTAPQRTSPYGHNGRPVTAGLVAANSGASAASSPTAAAKPTGEEK
                     ASAACETSSVAINATRPALHNASLPQAPTDGVLAAAVYQSEGEVHQSLERLESVITNT
                     SRVLKLLPDTIRRDHEQLLNLGLEAQMTELQQSRPTPQTQPRDTSSAKSSVFETYTLV
                     LIADSLSRNITKGVKRGVNEAIMLHLDHEVRHAIGNRLRQTQKNIIKSRLDEALKEST
                     TQFTAQLTQTVENLVKRELAEVLGSINGSLTSLVKENASLQKELNSIMSSGVLDEMRR
                     MREELCTLRESVAKRKATMPDSSLHATSSFQGRRSAPETILATALSMVREQQYRQGLE
                     YMLMAQQPSLLLRFLSILTRENENAYSELIENVETPNDVWCSVLLQLIEAAATEAEKE
                     VVVGVAIDILSERDQIAQNGALGSKLTTAMRAFERQARSETTSRSFLQCLKNLEKLLQ
                     S"
BASE COUNT      889 a    818 c    958 g    737 t
ORIGIN


  Query Match           100.0%;  Score 22;  DB 96;  Length 3402;
  Best Local Similarity  100.0%;  Pred. No. 3.5;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcgtcct 22
          ||||||||||||||||||||||
Db   1266 CAGCCGACGGTAGCTGCGTCCT 1287


RESULT   6
AC021683/c
LOCUS       AC021683   158566 bp    DNA             HTG       20-SEP-2000
DEFINITION  Homo sapiens clone RP11-13K12, WORKING DRAFT SEQUENCE, 14 unordered
            pieces.
ACCESSION   AC021683
VERSION     AC021683.3  GI:10198432
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 158566)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-13K12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 158566)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Sep 20, 2000 this sequence version replaced gi:7144863.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
```

```
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
        -------------- Project Information
            Center project name: L3343
            Center clone name: 13_K_12
        -------------- Summary Statistics
            Sequencing vector: M13; M77815; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 152322 bases at least Q40
            Consensus quality: 155539 bases at least Q30
            Consensus quality: 156788 bases at least Q20
            Insert size: 165000; agarose-fp
            Insert size: 157266; sum-of-contigs
            Quality coverage: 4.8 in Q20 bases; agarose-fp
            Quality coverage: 5.0 in Q20 bases; sum-of-contigs
        --------------.
        * NOTE: This is a 'working draft' sequence. It currently
        * consists of 14 contigs. The true order of the pieces
        * is not known and their order in this sequence record is
        * arbitrary. Gaps between the contigs are represented as
        * runs of N, but the exact sizes of the gaps are unknown.
        * This record will be updated with the finished sequence
        * as soon as it is available and the accession number will
        * be preserved.
        *        1     1431: contig of 1431 bp in length
        *     1432 1531: gap of      100 bp
        *     1532     2879: contig of 1348 bp in length
        *     2880 2979: gap of      100 bp
        *     2980     6254: contig of 3275 bp in length
        *     6255 6354: gap of      100 bp
        *     6355    11675: contig of 5321 bp in length
        *    11676 11775: gap of      100 bp
        *    11776    16766: contig of 4991 bp in length
        *    16767 16866: gap of      100 bp
        *    16867    22571: contig of 5705 bp in length
        *    22572 22671: gap of      100 bp
        *    22672    30550: contig of 7879 bp in length
        *    30551 30650: gap of      100 bp
        *    30651    38928: contig of 8278 bp in length
        *    38929 39028: gap of      100 bp
        *    39029    47882: contig of 8854 bp in length
        *    47883 47982: gap of      100 bp
        *    47983    60422: contig of 12440 bp in length
        *    60423 60522: gap of      100 bp
        *    60523    75477: contig of 14955 bp in length
        *    75478 75577: gap of      100 bp
        *    75578    92875: contig of 17298 bp in length
        *    92876 92975: gap of      100 bp
        *    92976   130172: contig of 37197 bp in length
        *   130173 130272: gap of      100 bp
        *   130273   158566: contig of 28294 bp in length.
FEATURES             Location/Qualifiers
     source          1. .158566
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-13K12"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1431
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    1532. .2879
                     /note="assembly_fragment"
     misc_feature    2980. .6254
                     /note="assembly_fragment"
     misc_feature    6355. .11675
                     /note="assembly_fragment"
     misc_feature    11776. .16766
                     /note="assembly_fragment"
     misc_feature    16867. .22571
                     /note="assembly_fragment"
     misc_feature    22672. .30550
                     /note="assembly_fragment"
     misc_feature    30651. .38928
                     /note="assembly_fragment"
     misc_feature    39029. .47882
                     /note="assembly_fragment"
     misc_feature    47983. .60422
                     /note="assembly_fragment"
     misc_feature    60523. .75477
                     /note="assembly_fragment"
     misc_feature    75578. .92875
                     /note="assembly_fragment"
     misc_feature    92976. .130172
                     /note="assembly_fragment"
     misc_feature    130273. .158566
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    40069 a  38517 c  37765 g  40914 t   1301 others
ORIGIN


  Query Match             80.9%;  Score 17.8;  DB 66;  Length 158566;
  Best Local Similarity   90.5%;  Pred. No. 1.1e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 cagccgacggtagctgcgtcc 21
           |||| ||||| ||||||||||
Db   61377 CAGCTGACGGAAGCTGCGTCC 61357


RESULT   7
AC005439
LOCUS       AC005439    29485 bp    DNA             INV       09-SEP-1998
DEFINITION  Drosophila melanogaster DNA sequence (P1 DS01989 (D353)), complete
            sequence.
ACCESSION   AC005439
VERSION     AC005439.1  GI:3548783
KEYWORDS    HTG.
SOURCE      Drosophila melanogaster (Subclones in sac from P1 clone DS01989
            (D353)) DNA.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 29485)
  AUTHORS   Celniker,S.E., George,R.A., Galle,R.F., Hoskins,R.A.,
            Svirskas,R.R., Harris,N.L., Agbayani,A., Arcaina,T.T., Baxter,E.,
            Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E.,
            Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,S.H., Lee,B., Lomotan,M.A., Mak,J., Mazda,P., Moshrefi,A.R.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,E.,
            Snir,E., Twomey,B., Wan,K.H., Zhang,R., Zieran,L.L. and Rubin,G.M.
  TITLE     Sequencing of Drosophila chromosome 2R, region 46F5-46F7
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 29485)
  AUTHORS   Celniker,S.E., George,R.A., Galle,R., Svirskas,R.R., Hoskins,R.A.,
            Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Chavez,C.,
            Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
            Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,
            Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S., Moshrefi,A.R.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,E.,
            Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R.,
            Zieran,L.L. and Kimmel,B.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-AUG-1998) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Sep 5, 1998 this sequence version replaced gi:3419743.
            Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
```

```
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://fruitfly.berkeley.edu/sequence/) or send
            email to bdgp@fruitfly.berkeley.edu.
            Library location: 69-21.
FEATURES              Location/Qualifiers
     source           1. .29485
                      /organism="Drosophila melanogaster"
                      /db_xref="taxon:7227"
                      /chromosome="2R"
                      /map="46F5-46F7"
                      /clone="P1 DS01989 (D353)"
BASE COUNT     8297 a   6045 c   6384 g   8759 t
ORIGIN


  Query Match             78.2%;  Score 17.2;  DB 4;  Length 29485;
  Best Local Similarity   86.4%;  Pred. No. 3.4e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 cagccgacggtagctgcgtcct 22
           |||| ||| | ||||||||||
Db   21681 CAGCTGACCGGAGCTGCGTCCT 21702


RESULT   8
AC017302/c
LOCUS       AC017302    33077 bp    DNA             HTG       09-DEC-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC017302
VERSION     AC017302.1  GI:6553684
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 33077)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10210213 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES              Location/Qualifiers
     source           1. .33077
                      /organism="Drosophila melanogaster"
                      /db_xref="taxon:7227"
BASE COUNT     9543 a   7246 c   7112 g   9176 t
ORIGIN


  Query Match             78.2%;  Score 17.2;  DB 64;  Length 33077;
  Best Local Similarity   86.4%;  Pred. No. 3.3e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 cagccgacggtagctgcgtcct 22
           |||| ||| | ||||||||||
Db   21449 CAGCTGACCGGAGCTGCGTCCT 21428


RESULT   9
HSJ858B16
LOCUS       HSJ858B16   36676 bp    DNA             PRI       12-DEC-1999
DEFINITION  Human DNA sequence from clone RP5-858B16 on chromosome 22 Contains
            the KIAA0542 gene and the PISD gene for phosphatidylserine
            decarboxylase (PSSC, EC 4.1.1.65) . Contains ESTs, GSSs and a
            putative CpG island, complete sequence.
ACCESSION   AL096768
VERSION     AL096768.7  GI:5596770
KEYWORDS    HTG; CpG island; KIAA0542; Phosphatidylserine Decarboxylase
            Proenzyme; PISD; PSSC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 36676)
  AUTHORS   Barlow,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 27, 1999 this sequence version replaced gi:5579004.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 22, constructed by the Sanger Centre Chromosome 22
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr22
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP5-858B16 is
            from the library RPCI-5 constructed at the Roswell Park Cancer
            Institute by the group of Pieter de Jong. For further details see
            http://bacpac.med.buffalo.edu/
            VECTOR: pCYPAC2
            IMPORTANT: This sequence is not the entire insert of clone
            RP5-858B16 It may be shorter because we only sequence overlapping
            sections once, or longer because we arrange for a small overlap
            between neighbouring submissions.
            The true left end of clone RP5-858B16 is at 1 in this sequence. The
            true left end of clone RP4-694E4 is at 36572 in this sequence.
FEATURES              Location/Qualifiers
     source           1. .36676
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="22"
                      /clone_lib="RPCI-5"
                      /clone="RP5-858B16"
     mRNA             join(<273. .454,1413. .1488,1978. .2073,2538. .2614,
                      3997. .4099,5606. .5702,8812. .8972,9435. .9509,10811. .10885,
                      11023. .11199,11271. .12948,14414. .14546,14669. .14747,
                      15784. .15895,15983. .16218)
                      /gene="dJ858B16.1"
                      /note="match: cDNAs: Em:AB011114"
                      /evidence=not_experimental
                      /product="dJ858B16.1 (KIAA0542 (isoform 2))"
     mRNA             join(<273. .454,1413. .1488,1978. .2073,2538. .2614,
                      3997. .4099,5606. .5702,8812. .8972,9435. .9509,10811. .10885,
                      11023. .11199,11271. .11524,12717. .12948,14414. .14546,
                      14669. .14747,15784. .15895,15983. .16217)
                      /gene="dJ858B16.1"
                      /note="match: ESTs: Em:Z78408 Em:R53362 Em:AI806913
                      Em:AA521233 Em:R53964 Em:AA158770 Em:AI858196 Em:AA594535
                      Em:AA836285 Em:AA974403 Em:AI425026 Em:AI809747
                      Em:AI978557 Em:AI141022 Em:AA159572 Em:AA553345
                      Em:AA582394 Em:R51102 Em:AA829254 Em:R54651 Em:AA971234
                      Em:AA846853 Em:N56896 Em:W90666 Em:AA118158 Em:AV030271
                      Em:AV116402 Em:AI024124 Em:AV122327 Em:N63934 Em:AI191287
```

```
                Em:AV031353 Em:AI623440 Em:AA779744 Em:AI126154
                Em:AA916775 Em:AW028290 Em:AI361548 Em:AI190893
                Em:AI019759 Em:N30173 Em:AI824779 Em:N64034 Em:AV075750
                Em:AW009329 Em:AV095355 Em:AA044074 Em:AA865102
                Em:AA398600 Em:AI232370 Em:AI770152 Em:AA637284
                Em:AA088219 Em:W25871 Em:AI961247 Em:AI094871 Em:AI192647
                Em:AA673371 Em:AW054842 Em:AI342072"
                /evidence=not_experimental
                /product="dJ858B16.1 (KIAA0542 (isoform 1))"
gene            join(273. .454,1413. .1488,1978. .2073,2538. .2614,3997. .4099,
                5606. .5702,8812. .8972,9435. .9509,10811. .10885,
                11023. .11199,11271. .12948,14414. .14546,14669. .14747,
                15784. .15895,15983. .16218)
                /gene="dJ858B16.1"
CDS             join(<273. .454,1413. .1488,1978. .2073,2538. .2614,
                3997. .4099,5606. .5702,8812. .8972,9435. .9509,10811. .10885,
                11023. .11199,11271. .11524,12717. .12948,14414. .14546,
                14669. .14747,15784. .15895,15983. .16087)
                /gene="dJ858B16.1"
                /note="different from published protein but supported by
                FGENES and GENSCAN"
                /codon_start=1
                /evidence=not_experimental
                /product="dJ858B16.1 (KIAA0542 (isoform 1))"
                /protein_id="CAB56392.1"
                /db_xref="GI:5921489"
                /translation="QAILHAERQLLYRSWFMWHQQAAARHQEQEWQTVACAHHRHGRL
                KKAFCLWRESAQGLRTERTGRVRAAEFHMAQLLRWAWSQWRECLALRGAERQKLMRAD
                LHHQHSVLHRALQAWVTYQGRVRSILREVAARESQHNRQLLRGALRRWKENTMARVDE
                AKKTFQASTHYRRTICSKVLVQWREAVSVQMYYRQQEDCAIWEAQKVLDRGCLRTWFQ
                RWWDCSRRSAQQRLQLERAVQHHHRQLLLEGLARWKTHHLQCVRKRLLHRQSTQLLAQ
                RLSRTCFRQWRQQLAARRQEQRATVRALWFWAFSLQAKVWATWLAFVLERRRKKARLQ
                WALQAYQGQLLQEGATRLLRFAASMKASRQQLQAQQQVQAAHSLHRAVRRCATLWKQK
                VLGRGGKPQPLAAIAPSRKVTFEGPLLNRIAAGAGDGTLETKRPQASRPLGALGRLAA
                EEPHALELPQKPQEHGLGMAQPAAPSLTRPFLAEAPTALVPHSPLPGALSSAPGPKQP
                PTASTGPELLLLPLSSFMPCGAAAPARVSAQRATPRDKPPVPSSLASVPDPHLLLPGD
                FSATRAGPGLSTAGSLDLEAELEEIQQQLLHYQTTKQNLWSCRRQASSLRRWLELNRE
                EPGPEDQEVEQQVQKELEQVEMQIQLLAEELQAQRQPIGACVARIQALRQALC"
CDS             join(<273. .454,1413. .1488,1978. .2073,2538. .2614,
                3997. .4099,5606. .5702,8812. .8972,9435. .9509,10811. .10885,
                11023. .11199,11271. .11528)
                /gene="dJ858B16.1"
                /note="original published form; could be artefactual as it
                is result of a large intron read-through
                match: proteins: Tr:O60289"
                /codon_start=1
                /evidence=not_experimental
                /product="dJ858B16.1 (KIAA0542 (isoform 2))"
                /protein_id="CAB56393.1"
                /db_xref="GI:5921490"
                /translation="QAILHAERQLLYRSWFMWHQQAAARHQEQEWQTVACAHHRHGRL
                KKAFCLWRESAQGLRTERTGRVRAAEFHMAQLLRWAWSQWRECLALRGAERQKLMRAD
                LHHQHSVLHRALQAWVTYQGRVRSILREVAARESQHNRQLLRGALRRWKENTMARVDE
                AKKTFQASTHYRRTICSKVLVQWREAVSVQMYYRQQEDCAIWEAQKVLDRGCLRTWFQ
                RWWDCSRRSAQQRLQLERAVQHHHRQLLLEGLARWKTHHLQCVRKRLLHRQSTQLLAQ
                RLSRTCFRQWRQQLAARRQEQRATVRALWFWAFSLQAKVWATWLAFVLERRRKKARLQ
                WALQAYQGQLLQEGATRLLRFAASMKASRQQLQAQQQVQAAHSLHRAVRRCATLWKQK
                VLGRGGKPQPLAAIAPSRKVTFEGPLLNRIAAGAGDGTLETKRPQASRPLGALGRLAA
                EEPHALEL"
repeat_region   620. .913
                /note="AluSx repeat: matches 1. .294 of consensus"
misc_feature    1444. .1687
                /note="match: GSS: Em:AQ545234"
repeat_region   3250. .3536
                /note="AluJo repeat: matches 2. .302 of consensus"
repeat_region   3634. .3836
                /note="L2 repeat: matches 2497. .2701 of consensus"
repeat_region   4356. .4646
                /note="AluSx repeat: matches 21. .312 of consensus"
repeat_region   4650. .4754
                /note="AluSg/x repeat: matches 200. .307 of consensus"
repeat_region   4990. .5281
                /note="AluSx repeat: matches 1. .292 of consensus"
repeat_region   5806. .6231
                /note="MLT1C repeat: matches 6. .460 of consensus"
repeat_region   6529. .6668
                /note="FRAM repeat: matches 1. .140 of consensus"
repeat_region   6671. .6966
                /note="AluY repeat: matches 1. .295 of consensus"
repeat_region   6985. .7081
                /note="LTR40a repeat: matches 1. .100 of consensus"
repeat_region   7929. .8097
                /note="MIR repeat: matches 82. .255 of consensus"
repeat_region   9766. .10068
                /note="AluY repeat: matches 1. .301 of consensus"
repeat_region   10069. .10096
                /note="14 copies 2 mer ta 100 conserved"
repeat_region   10104. .10399
                /note="AluJb repeat: matches 1. .293 of consensus"
repeat_region   12099. .12259
                /note="MIR repeat: matches 46. .208 of consensus"
repeat_region   13063. .13779
                /note="L2 repeat: matches 1570. .2750 of consensus"
misc_feature    complement(join(14467. .14747,15784. .16220))
                /note="match: STS: Em:AA521233"
misc_feature    complement(join(14667. .14747,15784. .16220))
                /note="match: STS: Em:R53964"
repeat_region   15108. .15404
                /note="AluSg repeat: matches 1. .295 of consensus"
mRNA            complement(join(16160. .17505,18222. .18382,18665. .18811,
                19003. .19141,19318. .19554,21353. .21518,23432. .23685,
                28473. .28494))
                /gene="PISD"
                /note="match: cDNAs: Em:M62722 Em:AL050371
                match: ESTs: Em:AA682112 Em:AA324939 Em:AI930241
                Em:AA166440 Em:AA166445 Em:AA118158 Em:AI019759
                Em:AL039237 Em:AA260704 Em:AI232370 Em:AA326672
                Em:AU035550 Em:AU051210 Em:AA297077 Em:AA144920
                Em:AA673371 Em:R00603 Em:AA298268 Em:AA237290 Em:AA615772
                Em:AA297754 Em:AA297076 Em:AA643609 Em:AV083921
                Em:AV122327 Em:AI094871 Em:AA717291 Em:H78465 Em:T96009
                Em:AA298523 Em:AV031353 Em:AI191287 Em:AA836285
                Em:AA916775 Em:AA297159 Em:AI361548 Em:AI806913 Em:R51102
                Em:N30173 Em:AA521233 Em:AA044074 Em:AI425026 Em:H69648
                Em:AA971234 Em:W00857 Em:AA553345 Em:H93993 Em:AA046738
                Em:AA637284 Em:AA151798 Em:W69452 Em:AA297086 Em:W69536
                Em:T10522 Em:AI024124 Em:H71693 Em:AA302951 Em:N98260
                Em:N63934 Em:AI978557 Em:R94341 Em:AA177542 Em:AA865102
                Em:N47783 Em:AI342072 Em:AA255719 Em:AA829254 Em:W69460
                Em:W69544 Em:AI014094 Em:AA779744 Em:AW028290 Em:N64034
                Em:AV075750 Em:H38339 Em:AV095355 Em:AI824779 Em:AA974403
                Em:AA908164 Em:AA088219 Em:AA298551 Em:N56896 Em:AV030271
```

```
  Query Match             78.2%;  Score 17.2;  DB 93;  Length 36676;
  Best Local Similarity   86.4%;  Pred. No. 3.2e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 cagccgacggtagctgcgtcct 22
           ||||| |||||||||| | ||
Db    1943 CAGCCCACGGTAGCTGCTTTCT 1964
```

```
RESULT  10
AC013925/c
LOCUS       AC013925    97308 bp    DNA             HTG       16-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC013925
VERSION     AC013925.1  GI:6437410
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 97308)
```

```
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10212147 by the submitter.
            For further information on this sequence e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .97308
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT    28683 a  21069 c  20380 g  27176 t
ORIGIN


  Query Match           78.2%;  Score 17.2;  DB 63;  Length 97308;
  Best Local Similarity  86.4%;  Pred. No. 2.4e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 cagccgacggtagctgcgtcct 22
           || ||||||||||||||| | |
Db   77336 CAACCGACGGTAGCTGCGACGT 77315


RESULT  11
AP001080/c
LOCUS       AP001080    151369 bp    DNA             PLN       12-AUG-2000
DEFINITION  Oryza sativa genomic DNA, chromosome 1, clone:P0499C11.
ACCESSION   AP001080 BA000010
VERSION     AP001080.1  GI:6815051
KEYWORDS    .
SOURCE      Oryza sativa (cultivar:Nipponbare) DNA, clone:P0499C11.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
            Oryza.
REFERENCE   1  (bases 1 to 151369)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
            clone:P0499C11
  JOURNAL   Published Only in DataBase (2000) In press
REFERENCE   2  (bases 1 to 151369)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JAN-2000) to the DDBJ/EMBL/GenBank databases. Takuji
            Sasaki, National Institute of Agrobiological Resources, Rice Genome
            Research Program; Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     The orientation of the sequence is from SP6 to T7 of the PAC clone.
            Genes were predicted from the integrated results of the
            following:GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as
            SplicePredictor (October1998 version). The genomic sequence was
            searched against the non-redundant database NRP(PIR,SWISSPROT,
            GENPEPT, PDB) from MAFF DNAbank and the cDNA sequence database at
            RGP. Protein similarities of the cording regions were searched
            against NRP with BLASTP2.0. ESTs represent the identified cDNA
            sequences using BLASTN2.0 with the corresponding DDBJ accession no.
            and RGP clone ID
            The sequence of this clone has an overlap with P0011D01 clone,
            DDBJ:AP000969 at the 3' end. This clone ends at the position 22,744
            of P0011D01. Detailed information on assemble quality together with
            annotation of this entry at
            http://www.dna.affrc.go.jp:82/genomicdata/GenomeFinished.html.
FEATURES             Location/Qualifiers
     source          1. .151369
                     /organism="Oryza sativa"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:4530"
                     /chromosome="1"
                     /clone="P0499C11"
     CDS             complement(join(1611. .1732,2015. .2089,2171. .2908,
                     2957. .3092,3875. .3967,4835. .4934,5232. .5349,5632. .5706,
                     5788. .5941))
                     /note="hypothetical protein"
                     /codon_start=1
                     /protein_id="BAA90339.1"
                     /db_xref="GI:6815052"
                     /translation="MATTDATIAHIMDEQEDIEIKYSKCWNPIRPPTTLLTSNGHRIC
                     IRPPFSAQSQEQGVASPIMGLWACNFIWDPGSGIGLICTGKINNPRRGVSIAKAQHNV
                     SYGCSRIVNVSPKSRRCHRVSIVVGTYENGIDSLPQFYFCSFQKFLSLYRKCLPRHRY
                     IQPSLGCLVESVYVTIPLALEVHNHEKEERDMKEPPVPLITLKVEAPPLSEEGIKGKL
                     NGAEINKAEQPLVEPTAEIPLSQIDLFAVPCDKEELCDNASLISMPQLVNEHAIPTVN
                     SYCTDFKHVVHIANEVEERELRSSLNTLGYVQFDDFCELDNFKEKLFAKSDLPYPTNV
                     IFHIFGEYNDLGIYFVHRVYICSDLEPPVHVDKICKLEKNVIANQIVSSLPCFDWKKQ
                     VVVNGLRKEHHMEKPRTVFREEGEDDVTMATTDATIAHIMDEQEDIEIKYSKCWNPIR
                     PPTTLLTSNGHRICIRPPFSAQSQEQGVASPIMGLWACNFNWDPGPGIGLICTGKINN
                     PRRGVSIAKAQHNVSYGCSRIVNVSPKS"
     CDS             complement(join(10030. .10270,10320. .10495,10995. .11141))
                     /note="hypothetical protein"
                     /codon_start=1
                     /protein_id="BAA90340.1"
                     /db_xref="GI:6815053"
                     /translation="MQQINGTRDEGFFTEVRKLAGFLIPVEASPTPPLNHEATARPLR
                     QEVGKKTSRYRAIIEPTQPGRSDSSVVAGQTVHGLVRPPSARSDRLRLWRLCIATKNQ
                     TSATDHPGQTDLRPTVRPQVTFQHNQPLVKRRYLLTRISEFGVLGLYGKLIQRAIQPM
                     NNPSNKHNFYQGEIAHTPKDTPPDITT"
     CDS             join(14385. .14517,14645. .14697,14798. .14914)
                     /note="EST AU078096(C53806) corresponds to a region of the
                     predicted gene.
                     hypothetical protein"
                     /codon_start=1
                     /protein_id="BAA90341.1"
                     /db_xref="GI:6815054"
                     /translation="MAQQQRRRRRGGSLAVLLLLLLLAALLLSLSLLARVDAAATVSS
                     ANLDWNEGEVAVATPLGQEAVAAAEEEGDRPPERVEMESINDYGQASANNRHNPHP"
     CDS             join(17001. .17084,17487. .17537,17828. .17879,18458. .18516,
                     19336. .19381,20159. .20283)
                     /note="EST AU081527(C53732) corresponds to a region of the
                     predicted gene.
                     hypothetical protein"
                     /codon_start=1
                     /protein_id="BAA90342.1"
                     /db_xref="GI:6815055"
                     /translation="MAGEASPPAPPEIPTRCHHCAGPLSKDMETSSWTVPPLVRDSFS
                     MIGSAIGGTAGAFYGFNHTMPFVRRYIKGPMWMHFLVGAPPVIVFSSACAGLAGGTIP
                     ALAQLVSSSYHSATASSFAHTTSRDEMHKARSSSTL"
     CDS             23322. .24899
                     /note="hypothetical protein"
                     /codon_start=1
                     /protein_id="BAA90343.1"
                     /db_xref="GI:6815056"
                     /translation="MDRYRVAPRPVYLSAAAAAGGDHGDGPRRLPRGVAGGGGAELDI
                     FSAERYFNADDVVKKEEYRDEEQPRQEVAVDAASQSGRTVACSSEASWTSRSGLLSGH
                     AQSAASAGAAAKLQMGGANGGGGAAAAAASAHRGKKPGSGQRWGLFSRDCPCAGRKAV
                     TVDVASEPRSPATPRTHARFENSHVVADSTIFKANAPPPPPPCEEEPVKMKISPGSST
                     FPPPLANSIFAAAPNRGGGGGGAAPFAAFPAPDIVGRRVVSSGGFTFPVAVGAAKVVS
                     TVGDEPPRESLEVFRPIDEESVLADPPADHLATVGGRGGGGGGLAAGFARAPGVAAVA
                     TDEEAMSDASSDLFDLESFAASSSFPTTCRGRSSRRNSREDDDDDEDLPYGAAAAAAA
                     AVEPALSECMYAPSEVSVVWSVATAEGGAFDAASVANFSSAASACCVEEFSFVPPPDS
                     AATAGGGGHEGFTAAMSRSAARKKGGGFLSSCRCEKAVSVGPTPVRMVRPEVNVKTTG
                     GGAARYHPGRVRMPVRT"
     CDS             complement(join(32830. .33189,33448. .33690,34214. .34312,
                     34814. .34993,35097. .35171))
                     /note="ESTs AU081555(E2011),C72657(E2011) correspond to a
                     region of the predicted gene.
                     Similar to putative phosphoglyceride transfer protein.
                     (AC002983)"
                     /codon_start=1
                     /protein_id="BAA90344.1"
                     /db_xref="GI:6815057"
                     /translation="MSFLFKSSNGATTEKNLTLEEQQEKINDLRKELGEHSSEEIQDF
                     LSDASCLRFLRARNWNVQKASKMMKSAVKWRVSYMPQKINWDDVAHEAETGKIYRADY
```

```
                KDKHGRTVLVLRPGLENTTSGKGQIKYLVYCLEKAIMSLTEDQEKMVWLTDFQSWTLG
                STPLKVTRETVNVLQDCYPERLGLAILYNPPRIFESFWKIVKPFLDHETYKKVKFVYS
                SDKESQKIMADVFDLDKLDSAFGGRNPATFEYNSYAERMRADDIKMGSSFKSNGITLP
                QDHSHVSADKEANGGGSDASSEASFYSGTDSPNHEEGGSILA"
CDS             join(40579. .40583,41045. .41324,42554. .42729,42883. .42889)
                /note="hypothetical protein"
                /codon_start=1
                /protein_id="BAA90345.1"
                /db_xref="GI:6815058"
                /translation="MSGARQRQAATGMASSSGTREVLMLEAPPAPSGAPWRAPPDAEA
                VDALPYIDGDYGDPAVKREVDRMVEEEMRRGSRKPADFLRDLPPVPAIGFENHPMLAK
                EYERVRAGKPPATIEMSRYGLEPPPANKRNDVAAWRQALRNAQSQLHHQIISF"
CDS             join(46218. .47573,47637. .47639)
                /note="Similar to geranylgeranyl hydrogenase. (AF069318)"
                /codon_start=1
                /protein_id="BAA90346.1"
                /db_xref="GI:6815059"
                /translation="MAAAVAAAACHSPARLVVTCSSSATPAPPRRPLRVAVVGGGPAG
                ASAAEALASAGAQAFLLERNPSGAKPCGGAIPLCMLDEFAIPRDLVDRRVTRMRVLSP
                SNLAADFSRALPPGAHIPMLRREVLDSFLRRRAADAGATLVPGLVTSLSLPADPATDP
                YLVHYISSDSGGPSSSRCVLEVDAVIGADGANSRVAREVGAGDYSTAIAFQERIRLPD
                EAMAYYDDLAEMYVGGDVSPDFYGWVFPKCDHVAVGTGTVAAKPEIKRLQSGIRARAA
                GKVAGGRVVRVEAHPIPEHPRPRRVVGRVALVGDAAGYVTRCSGEGIYFAARSGRLCG
                RAMADEWRLTGAVTEAGIRAGYLRRWDDEFRPTFRFLDLLQRVFYGGNAGREALVEMC
                ADEHVQRRTFDSYLHKRMAPAEPWGDLRLLWRTAAAMVRCGVLGREVERLRRLEQQAA
                DQ"
CDS             complement(join(48126. .48131,48532. .48613,49066. .49184,
                49984. .50103,50320. .50390,50808. .50955))
                /note="EST C26994(C50622) corresponds to a region of the
                predicted gene.
                Similar to ADP-ribosylation factor. (D17760)"
                /codon_start=1
                /protein_id="BAA90347.1"
                /db_xref="GI:6815060"
                /translation="MGLTFTKLFSRLFSKKEMRILMVGLDAAGKTTILYKLKLGEIVT
                TIPTIGFNVETVEYKNISFTVWDVGGQDKIRPLWRHYFQNTQGLIFVVDSNDRDRVVE
                ARDELHRMLNEDELRDAVLLVFANKQDLPNAMNAAEITDKLGLHSLRQRHWYIQSTCA
                TTGEGLYEGLDWLSSNIANKA"
CDS             complement(join(53616. .53715,53839. .53917,54090. .54259,
                54346. .54458,54525. .54593,54704. .54787,54874. .54930,
                55029. .55091,56173. .56229,56359. .56712))
                /note="ESTs AU068633(C30614),AU068634(C30614) correspond
                to a region of the predicted gene.
                Similar to Arabidopsis thaliana unknown protein.
                (AF049236)"
                /codon_start=1
                /protein_id="BAA90348.1"
                /db_xref="GI:6815061"
                /translation="MSHRRVDAAACDSWRPPPTPPPPHARRALALRVGGGGGGPAFAS
                LTVREGGGGGEEGEVGEGKGKQQQLPAAGVLVRHPLAMLALVPNAVALFAAGAAAGAV
                AKTITAPLDRVKLLMQTHSVRVVGESTKKGIGFLEAIAEIGKEEGLKGYWKGNLPQVI
                RIVPYSAVQLFSYEVYKKFFRRKDGELTVFGRLAAGACAGMTSTLVTYPLDVLRLRLA
                VQSGHSTMSQVAMNMLRDEGLASFYGGLGPSLIGIAPYIAVNFCVFDLMKKSVPEKYK
                SRPETSLATALLSATFATLMCYPLDTVRRQMQMKGSPYNTVLDAIPGIVERDGLIGLY
                RGFVPNALKNLPNSSIKLTAFDTVKTLISTGQKELEKLMQENQEKMS"
CDS             complement(join(57438. .57694,57993. .58332,58410. .58508,
                58590. .58935,58979. .59184,59277. .59468,59545. .59942,
                60037. .60307,60579. .60651,60870. .62380))
                /note="Similar to Arabidopsis thaliana chromosome II BAC
                F16P2 genomic sequence;mudrA like protein (AC004561)"
                /codon_start=1
                /protein_id="BAA90349.1"
                /db_xref="GI:6815062"
                /translation="MNWIGIGQIWHYRPTYGKAWRAKQVAMKVIWGDWDEAYVRLPTL
                MRAIKAKNPTMHFRVEAHPEKSRMVDGVQRRVFGRAYWIFGQSIEAFKHLRPVLAIDG
                TFLTGKYQGTLLTAIGVDAGLHLVPLAFALVEKENTSNWEWFINMLRNKLIGPNREVC
                IISDRHPGILNSIIHIMPHHLTIHHRWCMRHFCANFYTAGATTDQMKDLERICQINEK
                ALFLDEIKRLMGVVGERPKKWLEDHMPLKVKWARAFDTNGRRHSIMTSNMAESFNNVL
                RGIRKLPVTAIVAYTFSKCNSWFVDRHKEATVDILCGKKWPTKVKDMLEEQQRRTLGQ
                RAACFDFPSMKYEVSEQGGVTAAGVQWGGRHYVVVARDNTCSCQFPQLHHLPCSHMIT
                VCKLRGLDVEVAPRMCYEASNKAVQDSYSPRFEPYLDPSQWPSYDGEFFVPDLSLKNN
                TRGRRRTRRFKNDMDKAYKGSANRRSQKPNNTNSEPVMQNRCSLCHELGHKKTRMEHF
                PGLEDFYEEKHRAPQIASGERLKTLRVRGHTAHILFDDRYVPYLRRAKLLAFVTMAQR
                PVPLYNAAALTALVDRWRPETHTFHLPCGELTVTLEDVAMILGLPIRGQAVTGDTASG
                NWRERVEEYLGLEPPVAPDGQRQTKTSGVPLSWLRANFGQCPAEADEATVQRYCRAYV
                LYIFGSILFPDSGGDMASWMWLPLLADWDEAGTFSWGSAALAWLYRQLCDACRRQGGD
                ANLAGCVWLLQVWMWMRLPVGRPMWRTHQAWPHQDADRRPTVAHLWESVPSPVVGRRN
                LAYYHYTNEMDYLQPEHVVWMPYQAQEVLELELNPMCHIEDALKTLRCPLICFYAVEF
                QMCHRVMRQFGRLQTIPHRFSTSIDLHNHFNIYYIFELDILHTCRVDRRKNKKVTDWA
                YYHQDHITQWEKFEENGVPDQGQHNGTEFDLYLAWLHRTYRLVLRPAWTLADIADDPE
                DVEEQNEYDTRTRLGTTVETGPVRDRVARELLRTVNDAGVALGTAPGSEGEGGTLRNA
                LQRLRQRCRKLAARLGCRSTDVVEHAHAHQEGDEEGEVGDEEGEQDKEAEEGDEEEEG
                DEDREEEADELGPSQLEDAPESSQPDISQPGPSRPRRRRAPSQDWRYTPDVPRPRTRV

  Query Match          78.2%;  Score 17.2;  DB 13;  Length 151369;
  Best Local Similarity  86.4%;  Pred. No. 2.1e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 cagccgacggtagctgcgtcct 22
           ||  |||||| ||||||||||
Db 139950 CAAGCGACGGGAGCTGCGTCCT 139929


RESULT  12
AF081810/c
LOCUS       AF081810   161046 bp    DNA   circular  VRL       08-MAR-1999
DEFINITION  Lymantria dispar nucleopolyhedrovirus, complete genome.
ACCESSION   AF081810
VERSION     AF081810.1  GI:3822234
KEYWORDS    .
SOURCE      Lymantria dispar nucleopolyhedrovirus.
  ORGANISM  Lymantria dispar nucleopolyhedrovirus
            Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
            Nucleopolyhedrovirus.
REFERENCE   1  (bases 1 to 161046)
  AUTHORS   Kuzio,J., Pearson,M.N., Harwood,S.H., Funk,C.J., Evans,J.T.,
            Slavicek,J.M. and Rohrmann,G.F.
  TITLE     Sequence and analysis of the genome of a baculovirus pathogenic for
            Lymantria dispar
  JOURNAL   Virology 253 (1), 17-34 (1999)
  MEDLINE   99124785
REFERENCE   2  (bases 1 to 161046)
  AUTHORS   Kuzio,J., Pearson,M.N., Harwood,S.H., Funk,C.J., Evans,J.T.,
            Slavicek,J. and Rohrmann,G.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-1998) Department of Microbiology, Oregon State
            University, Corvallis, OR 97331-3804, USA
REFERENCE   3  (bases 1 to 161046)
  AUTHORS   Kuzio,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAR-1999) NCBI, 8600 Rockville Pike, Bethesda, MD
            20894, USA
FEATURES             Location/Qualifiers
     source          1. .161046
                     /organism="Lymantria dispar nucleopolyhedrovirus"
                     /db_xref="taxon:10449"
     repeat_region   39. .523
                     /function="enhancer; replication origin"
                     /rpt_family="hr1"
                     /rpt_type=dispersed
     CDS             673. .1410
                     /note="28772 Da; polyhedral matrix protein; Ld-PH;
                     LdOrf-1"
                     /codon_start=1
                     /product="polyhedrin"
                     /protein_id="AAC70186.1"
                     /db_xref="GI:3822235"
                     /translation="MHNFYNYSPALGKTYVYDNKYYKNLGTVIKQAKRQKHLEQHEIE
                     ERSLDHLDRYLVAEDPFYGPGKNQKLTLFKEIRNVKPDTMKLVVNWSGKEFLRETWTR
                     FMEDSFPIVNDQEVMDIYLTINVRPTRPNRCYKFVAQHALRCDEGYVPHEVIRIVEPS
                     TVENNEYRISLAKRGGGCPIRNLHSAYTTSFEHFLNSVIWDNFYKPIVYVGTTSAEEE
                     EILLEVSLVFKIKEFAPDAPLFQGPAY"
     CDS             complement(1418. .3085)
                     /note="61056 Da; LdOrf-2; Ld-Orf1629"
                     /codon_start=1
                     /product="essential structural protein pp78-81"
```

```
                     /protein_id="AAC70187.1"
                     /db_xref="GI:3822236"
                     /translation="MFKSYAGNEPESNLFHNEAYPPRDEALRLVSEVAAAAPASAQWL
                     IDKIMTSAVESPGVLTFDLRETVDLLRFFNRVCAAAAPVRSDDSPRIRRLQSMVDRLS
                     GHKAALQDLIARLYRTPDSLELYKEFLEAYRRHLVASRDERRIDSIFNEIVELRDAPS
                     PSPSPEPIFQKSSTSAPEPARQPAPAPARQPAPAPTRPIAPAPEPTRQETPIAPASEP
                     ARQESPIGSSAPEPIRQETPTGLFAPPPPPPPPPPPPPPEPLQQKSSAVPPPPPPLP
                     PPGAPDDPFGEIGQEVRPKPAAERAPTDALFAEIRRGVQLKPATERAPTYTPDALFAE
                     IRQGVKLKPAAERADEPPKPSSRAPLLLEIENRDKIKLKKVAPRATEPPASATNNPLM
                     QLLNKRLESMKMSSAEESDANYTSSSWSDAEDDSLLRDALRIKLALLGPRLSESERKR
                     IAKRLAGAKLKSAEKTLDELQARAIEPDNPLLSPPYQLTAPLYLHDLKLFESAVLDLF
                     NRGAYETALEKLEEALQVDLQAPSLQRMHDDISTYVKLQKKRLESEA"
CDS                  3087. .3911
                     /note="32438 Da; Ld-pk; LdOrf-3"
                     /codon_start=1
                     /product="protein kinase"
                     /protein_id="AAC70188.1"
                     /db_xref="GI:3822237"
                     /translation="MDALIGDFADFHKECSARTALHLVNGKFGKVSVWKHGPTQKSFF
                     YKRIEHKHFNAIEPFVHHLMKFNKYFLRLFYSLHSLREHLLVMDYIPDGDLFDLMQTE
                     PRLREPEISLIAYQLIDALQALHKHNVVHNDVKLENVLYRRFEQIYVCDYGLCKIAGS
                     PSTFEGTVDYFSPEKINKHAAAVHFDWWAVGVLLYEISTGKHPFKLDQDESLDVETLH
                     KRQIQLDVTFPADFDNPFLEEFICFLLGYCYDYRAHSYEVIQKNTYWKSIVHWKQR"
CDS                  complement(4501. .7590)
                     /note="109196 Da; LdOrf-4"
                     /codon_start=1
                     /product="mucin-like protein"
                     /protein_id="AAC70189.1"
                     /db_xref="GI:3822238"
                     /translation="MCKNCAFHVIIKNLKIAKSEVDFDSDGSENEDRLIVNRTGCRTD
                     AQPGPSRASDKRSSDSDRDSKRSKGSDSSSSDSEQEEPSRDPQPMLIPVEDKELNARR
                     KEILAKLNEERAEAIKMLRDQQAPPEVPRPQKSDSSSSSSSSSSSSSSGSDSDSDDECT
                     KNVSGKNIFNCYSRPEFEAKNVSEDLFEATGSEIAAEPAPEPFAEPADAEPAPEPFAE
                     PADAEPAPFDESALEEPVPSPSSCLSLSPGPMQFESVDEPATSPGEHKDPQPISNSKA
                     STTVSDKSRSPSADKHGSDRAKSPKSGCDRPKSSSAVSGKPKPEAKSKSSCDKPTGPG
                     VSTSGYDKPKSSSADKPSDKAKSSSTVKPKSSSADTPSDKAKSSSTVKPKSSSDKPSD
                     KAKSSSSSDKPCDKAKSSSTVKPKSSPDKPCDKAKSSSTVKPKSSSDKSPDKLSDKPS
                     DKAKSSSTVKPKSSPDKSPDKAKSSSTVKPKSSPDKSPDKAKSSSTVKPKSSPDKSPD
                     KAKSSSTVKPKSSSAVPDKPKHSSDKSRDATVKPKSSSAVPDKPSDKPSSDGSRSFRP
                     PSESGRDQPKSLAQTSPFLAAALARDMDERLNRDLEAAVASITGGVRPEFYPAPEPSA
                     PVDYFTLNSAEFAPEPAPEPVPEPSAPVDYFTLNSAEFAPEPAPEPAPEPAPEPAPEP
                     TSRFTSEPALGPVEPALEPVEPALEPVEPALGPGEPALEPGEPALEPGESTAEAASEL
                     AVERPAQPAPDLTSDSAVESTFGHKSTVASELAARPAAELAAAPAPELTPEHSEPAPQ
                     PASKPASEPSAHLTPDRTLPSPLTVKSLQSLMKSDDEEVVEDIFVKNEDEEDTVEVED
                     IYIKNENNELCLVGVPGDYEHKASSSTPSASAEVVDDEVEIMSFPENHKAKPGVFMMN
                     KRKKSDLMSVRKRPRYSQQPQPPPPPFPPQFPPYSFASPYAASASAHQYPPMYHMSPY
                     HLPQPSFVPVPQQFPRYVYAPTALPDHCYAPAEVPDHQEVLDDILTCLGPDTETPQN"
CDS                  7722. .8291
                     /note="21763 Da; LdOrf-5"
                     /codon_start=1
                     /product="LdOrf-5 peptide"
                     /protein_id="AAC70190.1"
                     /db_xref="GI:3822239"
                     /translation="MRTLVPHDRHHILFDAVDHLHDRVQGVVRVIIHAYFLGHVSVFI
                     QLEHAAFQVILTLDHHTVYVFLGLESLFGYDLFVIFYVLFYTKLDVLQHVQAHELGQI
                     VLQTLLHLSHTVLLHVLEHLVGHPVAESVLKEVDPYKLEVARVFQNGHHVKLLALVQG
                     NVYVYLFYYLIVVEADLLPLGILARSVYT"
CDS                  complement(8282. .8524)
                     /note="8960 Da; LdOrf-6"
                     /codon_start=1
                     /product="LdOrf-6 peptide"
                     /protein_id="AAC70191.1"
                     /db_xref="GI:3822240"
                     /translation="MSEPDGRRGRARGGAQTYRADGAARRLIARRLIVSIGRLIASRT
                     RSGLDGSIRYRYGKLISRTVRVDRAIVLKQFLTCQV"
CDS                  8619. .9194
                     /note="22146 Da; Ld-g22; LdOrf-7"
                     /codon_start=1
                     /product="LdOrf-7 peptide"
                     /protein_id="AAC70192.1"
                     /db_xref="GI:3822241"
                     /translation="MSLTRTNHVAQVVLLTRFDGRFSVTKTVNPCLFMFKDKNFRDVY
                     GQYVRFSKVSDSKTKVPKGYVLCKYQYMCVHDKLYGTRDFLKRFNVVPNFPGNVIVVQ
                     YMLVREDCVDACRRLEKDARPENEPYVLYGSEFPREHLSLSMTLDTKYCSYYDWSGTT
                     MSKLWAAAQRDLDKSKLAGSVNKTSNSIVCN"
repeat_region        9496. .10019
                     /function="enhancer; replication origin"
                     /rpt_family="hr2"
                     /rpt_type=dispersed
CDS                  complement(10115. .10474)
                     /note="13876 Da; LdOrf-8"
                     /codon_start=1
                     /product="LdOrf-8 peptide"
                     /protein_id="AAC70193.1"
                     /db_xref="GI:3822242"
                     /translation="MISSVDQTTLSASPRLSVTFAEKQLRSKNNFIYILFVSQKNKLI
                     ALIYCVLFEKQLYLHPVRIAEKQAYRVDLLRPVRKTSLSNRALAQADRRRRPTFFALQ
                     KSGLRIQDPYTEETLRS"
CDS                  10567. .11085
                     /note="19624 Da; LdOrf-9"
                     /codon_start=1
                     /product="LdOrf-9 peptide"
                     /protein_id="AAC70194.1"
                     /db_xref="GI:3822243"
                     /translation="MSLSSQTEAAARADKRKLQEDCSGVAKKLCILEAAQQAEAEPRP
                     TMWLVKDLDADIEYSDEEYEFDEEMEEGEGWDRRDLIFEIDPDFQPSAAPGPGTRIGK
                     YTTMFAGTHRKGDWQKCQFKQFQLIPDLKIQGLQTKVDAVMGSRKFKVEDVTLWLKDD
                     DRVFILVSVVLI"
CDS                  complement(11115. .11336)
                     /note="8543 Da; LdOrf-10; LdOrf-10"
                     /codon_start=1
                     /product="LdOrf-10 peptide"
                     /protein_id="AAC70195.1"
                     /db_xref="GI:3822244"
                     /translation="MILADWGAYQSITSKTKHCNLSNPRYQRLTSKTKHCNLNDSSYQ
                     RLTSKTKHCNLSNKPLLYPIIILHYKSFK"
CDS                  11530. .12264
                     /note="27768 Da; LdOrf-11; LdOrf-11"
                     /codon_start=1
                     /product="LdOrf-11 peptide"
                     /protein_id="AAC70196.1"
                     /db_xref="GI:3822245"
                     /translation="MSKRLCSASDSESDRPRGLYGDALIEALGLKPKESDRPGKRPRL
                     EILSKNVHGDKLTDVVKAFGAGNSELESSDTDLAMSSKRPEGNSELECSETTDLDLAM
                     PSGADYEKYSKSDLIKLLYITKTLYVLKKLDKKLLILKQKLAIQLKNKIVYSEALRKE
                     FAELQCHFRNNYSVRFPILMLPTLELQLRAIRDWCIATSAKLDAFIEDPVWRNGNSNI
                     HCFFIDVFIDTFCVYMQKAENLLFKD"
CDS                  12647. .13165
                     /note="20383 Da; LdOrf-12; LdOrf-12"
                     /codon_start=1
                     /product="LdOrf-12 peptide"
                     /protein_id="AAC70197.1"
                     /db_xref="GI:3822246"
                     /translation="MELYNDNIKKECLKNLIIDFVFERYLNLNEQIEKEFVEFVKQKN
                     INSSLVDTLMAHYATETLHESLSRFFEEEKNRIEQNRQRALLKRRIEQNRQRALQKQR
                     EHSKRKRRQEAPLQLKEDFPELAEEDFETLDALEAAAAAVSKDESPQEGESPQDVESI
                     QEGAKKMRLDEA"
CDS                  complement(13282. .13482)
                     /note="7643 Da; LdOrf-13; LdOrf-13"
                     /codon_start=1
                     /product="LdOrf-13 peptide"
                     /protein_id="AAC70198.1"
                     /db_xref="GI:3822247"
                     /translation="MILDILYFIKSGYISYTLAVYLFQCLIFLYLIKSGYISYTLAAK
                     CAIGLIVGFHKELKQNYVITIY"
CDS                  13685. .14755
                     /note="38636 Da; occluded; LdOrf-14; Ld-odv-e56"
                     /codon_start=1
                     /product="envelope protein"
                     /protein_id="AAC70199.1"
                     /db_xref="GI:3822248"
                     /translation="MSFFSSLRRVNKVYPNPNQFLAVDNLRVLNSTPNGFQNVFSAPS
                     TRQIGTNRFVPGYNLPNNRFVPSSDINRVMRNNDTAGMRNIFTSADNSQLNSLGQLRR
                     VDNIPDAGLHAAFLRRQGVKSNYPSTNTRTPGGVEGVLATNPRLHARLNGLKSVGTAL
                     LIGGGVYLVFSAATLVQDLVSALNRVGGSYYVVGSNGGDSSSVCLLTHRTCQLDTSNL
                     SDVSVCQFDPLLPDDPAGQRAICNGFNYEREQSVCRASDPYADPDSPQYVDISDLPAG
```

```
                HTLMCLEPYNLGDLIGDLGLDHLLGEEGLVDKSSNNSKDLSSKLWPLLVALIGIFAVV
                LILVFVFKRLMNKQTVDIRANA"
     CDS        complement(15221. .16921)

  Query Match          78.2%;  Score 17.2;  DB 58;  Length 161046;
  Best Local Similarity  86.4%;  Pred. No. 2.1e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcgtcct 22
          ||| ||||||| || |||||||
Db  31289 CAGTCGACGGTGGCGGCGTCCT 31268


RESULT  13
AP000969/c
LOCUS       AP000969   168655 bp    DNA             PLN       20-JAN-2001
DEFINITION  Oryza sativa genomic DNA, chromosome 1, clone:P0011D01.
ACCESSION   AP000969 BA000010
VERSION     AP000969.1  GI:6630680
KEYWORDS    .
SOURCE      Oryza sativa (cultivar:Nipponbare) DNA, clone:P0011D01.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
            Oryza.
REFERENCE   1  (bases 1 to 168655)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
            clone:P0011D01
  JOURNAL   Published Only in DataBase (1999) In press
REFERENCE   2  (bases 1 to 168655)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-DEC-1999) Takuji Sasaki, National Institute of
            Agrobiological Resources, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@abr.affrc.go.jp,
            URL:http://www.dna.affrc.go.jp:82/, Tel:81-298-38-7441,
            Fax:81-298-38-7468)
COMMENT     The orientation of the sequence is from SP6 to T7 of the PAC clone.
            Genes were predicted from the integrated results of the
            following:GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as
            SplicePredictor (October1998 version). The genomic sequence was
            searched against the non-redundant database NRP(PIR,SWISSPROT,
            GENPEPT, PDB) from MAFF DNAbank and the cDNA sequence database at
            RGP. Protein similarities of the cording regions were searched
            against NRP with BLASTP2.0. ESTs represent the identified cDNA
            sequences using BLASTN2.0 with the corresponding DDBJ accession no.
            and RGP clone ID.
FEATURES                Location/Qualifiers
     source             1. .168655
                        /organism="Oryza sativa"
                        /cultivar="Nipponbare"
                        /db_xref="taxon:4530"
                        /chromosome="1"
                        /clone="P0011D01"
     CDS                complement(join(6993. .7267,7613. .7664,7740. .8085,
                        8213. .8232))
                        /note="hypothetical protein"
                        /codon_start=1
                        /protein_id="BAA88527.1"
                        /db_xref="GI:6630681"
                        /translation="MQALESSHSFLSGLIQRVVPGRGVVVPELRRDETTGTSIGHARH
                        RKTTVLTVTSTRIRREGLRPHSTVERCTRPSPPPVVATTQLLSRSRSPLSPLRHVSGD
                        LSINPGGQRRCPFFLTELNQLHSLHGFKLVGNCSVLSLTPQADRSSARRMGRKPIADG
                        LKVDFFHNWMWMHARRHRGIAHRNRPSSRQKAQSGTRFLLSLSLSLSFSIVDHACLFF
                        LDLARAHAYALG"
     CDS                join(10822. .10871,11189. .11415,13238. .13635)
                        /note="EST D22716(C1010) corresponds to a region of the
                        predicted gene.
                        hypothetical protein"
                        /codon_start=1
                        /protein_id="BAA88528.1"
                        /db_xref="GI:6630682"
                        /translation="MPVTEACTFTFQKETEGFSPRGILPRRRLGTTSSYNGAELASSA
                        SEHGSSPSPRERTQLPSLDSTGGFFLHKRHHKDGAKRPAVRALIPLPFQFPIHPNPKS
                        EVATAESPNSCVQFLACLVFFSPHSDRSRCCVRFFELWSMASIVLLLSELLGGESTSV
                        MAADWYMSGHSLREFRPVAAAPAAAAAVAKCERPAAEAAGEKKKEESFEDLAAVSRIA
                        VDVMWP"
     CDS                complement(join(14110. .14217,14334. .14417,15213. .15374,
                        15966. .16286,16374. .16598,16697. .16726))
                        /note="EST C25991(C11435) corresponds to a region of the
                        predicted gene.
                        Similar to similar to PHZF, catalyzing the hydroxylation
                        of phenazine-1-carboxylic acid to
                        2-hydroxy-phenazine-1-carboxylic acid. (AC004044)"
                        /codon_start=1
                        /protein_id="BAA88529.1"
                        /db_xref="GI:6630683"
                        /translation="MVKNAIQYAVVDAFAAEPFKGNPAAVCLLEGEDAAAAADERWMQ
                        SVAAEFNLSETAFLIRDPSSSAAAADAAPRFRLRWFTPVAEVNLCGHATLASAHFLFT
                        TVLAKQQHAAAAMVEFVTRSGILTAKKVPAPPPPANDGGVPGEEKLFIELDFPMIDLV
                        EYDSAETLSIPETLNGARVVSVWKSSTAGDLIVELSSGKEVADIIPNINEIKKCDGRG
                        VIVTGPAPAGSDYDFFSRFFCPKFGIDEDPVCGSAHCVLAPYWGGKLGKQKLTAFQAS
                        PRSGTLYLELDGENRRVRIQGEAVTVMAGTLLA"
     CDS                complement(join(17773. .17878,18808. .18902,19465. .19623,
                        20350. .20379,20482. .20688))
                        /note="ESTs AU070430(S15424), AU082086(E11580) correspond
                        to a region of the predicted gene.
                        Similar to MF1. (AB011804)"
                        /codon_start=1
                        /protein_id="BAA88530.1"
                        /db_xref="GI:6630684"
                        /translation="MASALLRKATVGGSAAAAAARWASRGLASVGSGSDIVSAAPGVS
                        LQKARSWDEGVATNFSTTPLKDIFHGKKVVIFGLPGAYTGVCSQAHVPSYKNNIDKLK
                        AKGVDSVICVSVNDPYALNGWAEKLQAKDAIEFYGDFDGSFHKSLDLEVDLSAALLGR
                        RSHRWSAFVDDGKIKAFNVEVAPSDFKVSGAEVILDQI"
     LTR                complement(23379. .24781)
                        /note="3' LTR"
     CDS                complement(join(25866. .28616,28907. .28978))
                        /note="Similar to copia-type pol polyprotein. (AF105716)"
                        /codon_start=1
                        /protein_id="BAA88531.1"
                        /db_xref="GI:6630685"
                        /translation="MFSSLDPNGTSQENIIFGDDGKGKASPCWYEESSQTSKARHILG
                        LTNIQFEKDRVCSACQAGKQIGAHHPVKNVMTTTRPLELLHMDLFGPIAYLSIGGNKY
                        GLVIVDDFSCFTWVFFLHDKSETQAIFKKFARRAQNEFDLKIKNIRSDNVKEFKNTCI
                        ESFLDEEGIKHEFSAPYSPQQNGVAERKNRTLIEIARTMLDEYKTSDRFWAEAVNTVC
                        HDINRLYLHRLLKKTPYELLTGNKPNVSYFRVFGSKCYILNKKARSSKFAPKVDGGFL
                        LGYGSNECAYRVFNKTSGIVEIAPDVTFDETNGSQVEQVDSHVLGEEEDPREAIKRLA
                        LGDVRPREPQQGASSSTQVEPPTSTQANDPSTSSLDQGEEGEQVPPSSINLAHPRIHQ
                        SIQRDHPTDNILGDINKGVSTRSHIANFCEHYSFVSSLEPLRVEEALNDPDWVMAMQE
                        ELNNFTRNEVWTLVERSRQNVIGTKWIFRNKQDEAGVVIRNKARLVAQGFTQIEGLDF
                        GETFAPVARLESIRILLTFATNLNFKLYQMDVKSAFLNGLINELVYVEQPPGFKDPKY
                        PNHVYKLHKALYELKQAPRAWYECLRNFLVKNGFEIGKADSTLFTKRHDNDIFVCQIY
                        VDDIIFGSTNKSFSEEFSRMMTKRFEMSMMGELKFFLGLQIKQLKEGTFICQTKYLKD
                        MLKKFGMENAKPIHTPMPSNGHLDLNEQGKDVDQKVYRSIIGSLLYLCASRPDIMLSV
                        CMCARFQAAPKECHLVAVKRILRYLVHTPNLGLWYPKGARFDLIGYADADYAGCKVDR
                        KSTSGTCQFLGRSLVSWSSKKQNSVALSTAEAEYISTGSCCAQLIWMKQTLRDYGLNV
                        SKIPLLCDNESAIKIANNPVQHSRTKHIDIRHHFLRDHSTRGDIDIQHVRIDKQLADI
                        FTKPLDEARFCELRSELNILDSRNVA"
     LTR                complement(31265. .32660)
                        /note="5' LTR"
     CDS                join(46781. .46827,47917. .48214,50464. .50528,50674. .50765,
                        51357. .51460)
                        /note="hypothetical protein"
                        /codon_start=1
                        /protein_id="BAA88532.1"
                        /db_xref="GI:6630686"
                        /translation="MQNDIVSLAGLHAHETCQWAPYVCALALPRKQPARVENGRSDGH
                        RRGAAISGQAEAMSPCPHACDMSPRPTSHPAEEETTRKRTTARQNSTPVTYTVRSALM
                        HVIAPSQVQNHRQRYNWHTDITTRIDIAATPKDECGRRRIRRLRGACGDFVNLKGLPA
                        QSSKMSIGCHTCHGTHAAASCRGVPAEEETTSRIPHPHAVQ"
     CDS                complement(join(55289. .55492,56957. .57104,57426. .57596,
                        57653. .57699))
                        /note="hypothetical protein"
```

```
                     /codon_start=1
                     /protein_id="BAA88533.1"
                     /db_xref="GI:6630687"
                     /translation="MARPATAWITATPRRRRPTTAALCQQIEDGSMGVQIRTRRSRCE
                     ERETAARTTAAVAAGDVGTGGHASEKAGIHRSHNTTLVCWTWTWTVESEVGSHGGEWL
                     CRTEGRLNLVISSAGESVVMQKGDAEVTEETQWEWTNLDTEEEAMCRIWYGNQNSQCE
                     RCIKKKKITIAASTYATAKNKQQDILTLA"
     CDS             join(61440. .63104,63219. .63302,63477. .63581,63692. .63800,
                     64338. .64489,64592. .64738)
                     /note="Similar to Arabidopsis thaliana DNA chromosome 4,
                     BAC clone T5L19 (AL049481)"
                     /codon_start=1
                     /protein_id="BAA88534.1"
                     /db_xref="GI:6630688"
                     /translation="MDSRSPKIPRRSPDSKDKDSDRNKERDAKNDWDSSRAYGSETDC
                     KEEMCDSNKRKGLTMGEIVVDNSRSVDSCHETELHVLRDDRQDKSVEIKDILHDGVAK
                     SDYAQRQIDLDSERRNGTGDNSRVDVLRDDKLDSGRDRNWSDRTREPEGSKDYVRNRQ
                     WQDSKEANDSEWKNAHERLDGGSFHGRAGYRRDSRGRSESIRGSSTYGGRYDSSDSIE
                     IRPNNNLDFGREGSVSGRRYDVGAHRDATPGTNGDKSANPEADQSGSTSTISQFPQHG
                     PKGDRSSRGRGRPNSRDSQRVGGTLPIMPPPFGPLGLPPGPMQHIGPNIPHSPGPLLP
                     GVFVPPFPGPLVWPGARGVDVNMLSVPPNLPIPPVAAEHRFAPSMGAGPGHNIHLNQI
                     GSGIGAPTNVSGLSFHQLGTQSREMAHDKPPAGGGWTPHRNSGPTRKAPSRGEQNDYS
                     QNFVDTGMRPQNFIRELDLTSVAEDYPKLRELIQRKDEIVANSASPPMYYKCDLRQHV
                     LSPEFFGTKFDVILVDPPWEEYVHRAPGITDHIEYWNAEEIMNLKIEAIADTPSFVFL
                     WVGDGVGLEQGRQCLKKWGFRRCEDVCWVKTNKKNATPSLRHDSHTILQHSKEHCLMG
                     IKGTVRRSTDGHVIHANIDTDIIIAEEPTDGSTKKPEDMYRIIEHFALGKRRLELFGE
                     DHNIRPGWLTLGKGLSYSNFNKEAYIKNFADKDGKVWQGGGGRNPPPEAPHLVVTTPE
                     IEGLRPKSPPQKN"
     CDS             join(65836. .65962,66040. .66217,67638. .67719,68289. .68465,
                     68556. .68723,68848. .69039)
                     /note="ESTs D39011(R0609), AU032023(R3215) correspond to a
                     region of the predicted gene.
                     Similar to 26S proteasome, non-ATPase subunit. (Y13071)"
                     /codon_start=1
                     /protein_id="BAA88535.1"
                     /db_xref="GI:6630689"
                     /translation="MERLQRIFGASGMGQPPTDSPLLDSSEQVYISSLALLKMLKHGR
                     AGVPMEVMGLMLGEFVDDYTVRVVDVFAMPQSGTGVSVEAVDHVFQTNMLDMLKQTGR
                     PEMVVGWYHSHPGFGCWLSGVDINTQQSFEALNPRAVAVVIDPIQSVKGKVVIDAFRL
                     INPQTMMLGQEPRQTTSNVGHLNKPSIQALIHGLNRHYYSIAINYRKNELEEKMLLNL
                     HKKKWTDGLILKSFDTHSKTNEQTVQEMLNLAIKYNKAVQEEDELPPEKLAIANVGRQ
                     DAKKHLEEHVSNLMSSNIVQTLGTMLDTVVF"
     CDS             71085. .72281
                     /note="Similar to serpin - wheat. (S65782)"
                     /codon_start=1
                     /protein_id="BAA88536.1"
                     /db_xref="GI:6630690"
                     /translation="MAPPAMAEEAKVSCLPLAREVGRRAAAAGGGQGRNFIVSPLSFH
                     AALALVADGARGETQRELLGFLGSPSLAELHRSPTTRLVARLRHLPNTSFACGVWVDR
                     GRALTPEFADAAASRYAAVAEPADFATQPEQARERVNAFVSDATEGLIRDVLPPNSVD
                     SSTVVVLANAVHFKGTWSLPFHPSATFHAPFHLLDGGAVRAPFMTTEIPFERHVAAFP
                     GFTALKLPYKNVGGGGGDGVPRAAFYMLLLLPDGDGALKLADLYDMAVTTPEFIKKH
                     TPAAEAPVRRLMVPKFKFSFKFEAKSDMRKLGVTRAFAGGDFSGMVTGGDGLFIAEVY
                     HQATIEVDELGTVAAASTAVVMMQKGSSLPPVDFVADRPFLFAVVEELTGAVLFLGHV
                     VNPLAE"
     CDS             complement(join(72911. .73060,73798. .74024,74469. .74535))
                     /note="ESTs AU082078(C53435), C28061(C53848) correspond to
                     a region of the predicted gene.
                     Similar to PGP224. (AF049928)"
                     /codon_start=1
                     /protein_id="BAA88537.1"
                     /db_xref="GI:6630691"
                     /translation="MIHRAVCYSLTCDSLISEGKIKGFIGATCPCFLFGRNAQFLGSG
                     TLAGSCTTHCMLWGLLTSLCCVFTGGLVLAVPGSAVACYACGYRQALRAKYNLPEAPC
                     GDLTTHLFCHLCAICQEYREIRERTDSGSSAPTVTPPAIQTMDEP"
     CDS             complement(join(81876. .81979,82580. .82835,83032. .83196,
                     83667. .83795,83939. .84043,84886. .85016,85649. .85742))
                     /note="hypothetical protein"
                     /codon_start=1

Query Match              78.2%;  Score 17.2;  DB 13;  Length 168655;
Best Local Similarity    86.4%;  Pred. No. 2e+02;
Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 cagccgacggtagctgcgtcct 22
           ||  |||||| ||||||||||
Db   11325 CAAGCGACGGGAGCTGCGTCCT 11304


RESULT  14
AC007417
LOCUS       AC007417   175682 bp    DNA             INV       02-MAR-2001
DEFINITION  Drosophila melanogaster, chromosome 2R, region 47A-47B, BAC clone
            BACR48F07, complete sequence.
ACCESSION   AC007417
VERSION     AC007417.4  GI:13184056
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 175682)
  AUTHORS   Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
            Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
            Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
            Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
            Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
            Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
            McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
            Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
            Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
            Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
            Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE     Sequencing of Drosophila chromosome 2R, region 47A-47B
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 175682)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-APR-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Mar 2, 2001 this sequence version replaced gi:5670592.
            Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            This sequence was assembled using end sequences from a whole genome
            shotgun and from subclones of this BAC and its neighboring clones.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu.
FEATURES              Location/Qualifiers
     source           1. .175682
                      /organism="Drosophila melanogaster"
                      /strain="y; cn bw sp"
                      /db_xref="taxon:7227"
                      /chromosome="2R"
                      /map="47A-47B"
                      /clone="BACR48F07 (D625)"
                      /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                      Drosophila melanogaster BAC library, partial EcoRI in
                      pBACe3.6)"
BASE COUNT    49836 a  38215 c  38034 g  49597 t
ORIGIN
```

```
  Query Match             78.2%;  Score 17.2;  DB 4;  Length 175682;
  Best Local Similarity   86.4%;  Pred. No. 2e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcgtcct 22
          |||| ||| | ||||||||||
Db  35210 CAGCTGACCGGAGCTGCGTCCT 35231


RESULT  15
AE003830
LOCUS       AE003830   261846 bp    DNA             INV       04-OCT-2000
DEFINITION  Drosophila melanogaster genomic scaffold 142000013386047 section 13
            of 52, complete sequence.
ACCESSION   AE003830 AE002787
VERSION     AE003830.2  GI:10727666
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 261846)
  AUTHORS   Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
            Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
            Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
            Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
            Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
            Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
            Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
            Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
            Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
            Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
            Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
            Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
            Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
            Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
            Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
            Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
            Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
            Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
            Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
  TITLE     The genome sequence of Drosophila melanogaster
  JOURNAL   Science 287 (5461), 2185-2195 (2000)
  MEDLINE   20196006
REFERENCE   2  (bases 1 to 261846)
  AUTHORS   Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     On Oct 9, 2000 this sequence version replaced gi:7303755.
FEATURES             Location/Qualifiers
     source          1. .261846
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="2R"
     mRNA            join(960. .1069,1135. .1237,1290. .1357,1416. .1731,
                     2658. .2759,3801. .4142,4200. .4354,4727. .4835,6101. .6289,
                     6499. .6941,6999. .7841,7921. .8315)
                     /gene="Pfk"
                     /product="CT13302"
                     /db_xref="FLYBASE:FBan0004001"
                     /db_xref="FLYBASE:FBgn0003071"
     gene            <960. .>8315
                     /gene="Pfk"
                     /note="CG4001"
                     /map="46E3-46E6"
                     /db_xref="FLYBASE:FBan0004001"
                     /db_xref="FLYBASE:FBgn0003071"
     CDS             join(1026. .1069,1135. .1237,1290. .1357,1416. .1731,
                     2658. .2759,3801. .4142,4200. .4354,4727. .4835,6101. .6289,
                     6499. .6941,6999. .7841,7921. .8059)
                     /gene="Pfk"
                     /note="Pfk gene product [alt 1]"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0004001"
                     /db_xref="FLYBASE:FBgn0003071"
                     /protein_id="AAF58840.1"
                     /db_xref="GI:7303792"
                     /translation="MHSIKFRVFTKLKPIFLEINGRIPICRHFHGPTTFRLEISNKTP
                     PIRQKLTFPNIGIQCTRSHHLCCPRDISGNTLLSVKFNCKRHCIKLRSDSGDQKNDSP
                     GEKNIQKDKSAQRCGKPINNLHNGFLNAVNYSEKNAVKKKKSAPKRKCGKSVDELRKC
                     LRTMQDVIDFVHPVKPFKDKGLAVFTSGGDSQGMNAAVRACVRMAIYLGCKVYFIREG
                     YQGMVDGGDCIQEANWASVSSIIHRGGTIIGSARCQDFRERQGRLKAANNLIQRGITN
                     LVVIGGDGSLTGANLFRQEWSSLLDELVKNKTITTEQQEKFNVLHIVGLVGSIDNDFC
                     GTDMTIGTDTALHRIIEAIDAISSTAYSHQRTFIMEVMGRHCGYLALVGGLACEADFI
                     FIPEMPPKVDWPDRLCSQLAQERSAGQRLNIVIVAEGAMDREGHPITAEDVKKVIDER
                     LKHDARITVLGHVQRGGNPSAFDRILACRMGAEATLALMEATKDSVPVVISLDGNQAV
                     RVPLMECVERTQAVAKAMAEKRWADAVKLRGRSFERNLETYKMLTRLKPPKENFDADG
                     KGIEGYRLAVMHIGAPACGMNAAVRSFVRNAIYRGDVVYGINDGVEGLIAGNVRELGW
                     SDVSGWVGQGGAYLGTKRTLPEGKFKEIAARLKEFKIQGLLIIGGFESYHAAGQIADQ
                     RDNYPQFCIPIVVIPSTISNNVPGTEFSLGCDTGLNEITEICDRIRQSAQGTKRRVFV
                     IETMGGYCGYLATLAGLAGGADAAYIYEEKFSIKDLQQDVYHMASKMAEGVSRGLILR
                     NEKASENYSTDFIYRLYSEEGKGLFTCRMNILGHMQQGGSPTPFDRNMGTKMAAKCVD
                     WLAAQIKANIDANGVVNCKSPDTATLLGIVSRQYRFSPLVDLIAETNFDQRIPKKQWW
                     LRLRPLLRILAKHDSAYEEEGMYITVEEECDTDAVA"
     mRNA            join(2152. .2759,3801. .4142,4200. .4354,4727. .4835,
                     6101. .6289,6499. .6941,6999. .7841,7921. .8315)
                     /gene="Pfk"
                     /product="CT13275"
                     /db_xref="FLYBASE:FBan0004001"
                     /db_xref="FLYBASE:FBgn0003071"
     CDS             join(2613. .2759,3801. .4142,4200. .4354,4727. .4835,
                     6101. .6289,6499. .6941,6999. .7841,7921. .8059)
                     /gene="Pfk"
                     /note="Pfk gene product [alt 2]"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0004001"
                     /db_xref="FLYBASE:FBgn0003071"
                     /protein_id="AAF58841.1"
                     /db_xref="GI:7303793"
                     /translation="MNSEINQRFLARGSQKDKGLAVFTSGGDSQGMNAAVRACVRMAI
                     YLGCKVYFIREGYQGMVDGGDCIQEANWASVSSIIHRGGTIIGSARCQDFRERQGRLK
                     AANNLIQRGITNLVVIGGDGSLTGANLFRQEWSSLLDELVKNKTITTEQQEKFNVLHI
                     VGLVGSIDNDFCGTDMTIGTDTALHRIIEAIDAISSTAYSHQRTFIMEVMGRHCGYLA
                     LVGGLACEADFIFIPEMPPKVDWPDRLCSQLAQERSAGQRLNIVIVAEGAMDREGHPI
                     TAEDVKKVIDERLKHDARITVLGHVQRGGNPSAFDRILACRMGAEATLALMEATKDSV
                     PVVISLDGNQAVRVPLMECVERTQAVAKAMAEKRWADAVKLRGRSFERNLETYKMLTR
                     LKPPKENFDADGKGIEGYRLAVMHIGAPACGMNAAVRSFVRNAIYRGDVVYGINDGVE
                     GLIAGNVRELGWSDVSGWVGQGGAYLGTKRTLPEGKFKEIAARLKEFKIQGLLIIGGF
                     ESYHAAGQIADQRDNYPQFCIPIVVIPSTISNNVPGTEFSLGCDTGLNEITEICDRIR
                     QSAQGTKRRVFVIETMGGYCGYLATLAGLAGGADAAYIYEEKFSIKDLQQDVYHMASK
                     MAEGVSRGLILRNEKASENYSTDFIYRLYSEEGKGLFTCRMNILGHMQQGGSPTPFDR
                     NMGTKMAAKCVDWLAAQIKANIDANGVVNCKSPDTATLLGIVSRQYRFSPLVDLIAET
                     NFDQRIPKKQWWLRLRPLLRILAKHDSAYEEEGMYITVEEECDTDAVA"
     mRNA            join(8915. .9052,9216. .9726,10663. .10911,10976. .11174,
```

```
                11235. .11542,11596. .>12210)
                /gene="CG11866"
                /product="CT37014"
                /db_xref="FLYBASE:FBan0011866"
                /db_xref="FLYBASE:FBgn0033486"
     gene       <8915. .>12210
                /gene="CG11866"
                /map="46E4-46E6"
                /db_xref="FLYBASE:FBan0011866"
                /db_xref="FLYBASE:FBgn0033486"
     CDS        join(9286. .9726,10663. .10911,10976. .11174,11235. .11542,
                11596. .12210)
                /gene="CG11866"
                /note="CG11866 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0011866"
                /db_xref="FLYBASE:FBgn0033486"
                /protein_id="AAF58839.1"
                /db_xref="GI:7303791"
                /translation="MVSTRQMCSGGGTSASEGVSPFGGAGSGSASGTGMASGSAVSSG
                SGTGVQEPPESTANRSSVLRRTTSYQGNQAHSSYRVVSAGSLAESASAVTRLHSYTMQ
                MQRQRPPTATTTTTSAPSLYDLPNELIEKILSYVDYKKVSNLRLVSHRMNDICMAML
                NTAFTKQIKTTLSRFQAIKASMPRRESHRRNHPLACECDIIETCYMRLSLLQMSMGKH
                IERGHCCFFPGAILDEVQAILNYISITPRLQRPYRVTDELFDLSTMAMEYFKDRIEAT
                LPGLAYFNKDFYTLPTTTKRPTLAISSDLEDSASNSPPQNHMVLRKGIRKIKQGMKMY
                NNQLSVLRTELRTCKRKAAEQSKQLAEQSNLISEQQKQTLEYANRLDENDKKNEEMSR
                KFSTLLQELNKCKTELQFWRSKSPAIPAVCSSCNQKVAPVVPPEDFQALVNQGVDPEH
                FIIINEDVDAESDVSVGELKEFASPDESTTAKLLAVSTAAKNLKRKMPSESQSNAIAS
                TSKAAEVAQSQSQPMATGNAGAAKAAGGYSPKLFYGNHQRSGVIVSPVSQNLGVAATT
                DNVAPGSEALEEPEEAKKARRVQKANRYVNTPGKRSK"
     mRNA       complement(join(<12115. .12937,12999. .13642,13694. .13708))
                /gene="CG17753"
                /product="CT11457"
                /db_xref="FLYBASE:FBan0017753"
                /db_xref="FLYBASE:FBgn0033487"
     gene       complement(<12115. .>13708)
                /gene="CG17753"
                /map="46E4-46E6"
                /db_xref="FLYBASE:FBan0017753"
                /db_xref="FLYBASE:FBgn0033487"
     CDS        complement(join(12802. .12937,12999. .13642,13694. .13708))
                /gene="CG17753"
                /note="CG17753 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0017753"
                /db_xref="FLYBASE:FBgn0033487"
                /protein_id="AAF58838.1"
                /db_xref="GI:7303790"
                /translation="MSSIKIEFAVQMRRGDESYAGALRSALDGVGQVEIDTQEGRVII
                QTQRPWSEIQDKIEATGVRAVLSGFGGQSAVALINTTGSVVDKTPIQGVVRFTTITAD
                KKPGVVVDGVVDGLSPGLHGLHIHESGDTSAGCSSVGEHYNPRQSPHGSPAAGAEERH
                AGDLGNIRADENGRATFRFVDPVLEVWDIIGRAVVLTANADDLGRGGNDQSLIDGNSG
                ERIACGIIARSAGILENFKRICACDGVTLWDERNKPLAGKERSQKL"
     mRNA       join(<18636. .18761,19765. .19876,19968. .20213,22806. .22878,
                22958. .23344,23489. .23633,23934. .24407,25026. .>25202)
                /gene="CG11867"
                /product="CT37016"
                /db_xref="FLYBASE:FBan0011867"
                /db_xref="FLYBASE:FBgn0033488"
     gene       <18636. .>25202
                /gene="CG11867"
                /map="46E4-46E8"
                /db_xref="FLYBASE:FBan0011867"
                /db_xref="FLYBASE:FBgn0033488"
     CDS        join(18636. .18761,19765. .19876,19968. .20213,22806. .22878,
                22958. .23344,23489. .23633,23934. .24407,25026. .25202)

  Query Match             78.2%;  Score 17.2;  DB 5;  Length 261846;
  Best Local Similarity   86.4%;  Pred. No. 1.8e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 cagccgacggtagctgcgtcct 22
           |||| ||| | ||||||||||||
Db  125602 CAGCTGACCGGAGCTGCGTCCT 125623


Search completed: September 21, 2001, 00:35:41
Job time: 29596 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: September 21, 2001, 02:42:38 ; Search time 330.77 Seconds
(without alignments)
41.763 Million cell updates/sec

Title: US-09-138-735-9
Perfect score: 22
Sequence: 1 cagccgacggtagctgcgtcct 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0601:*
1: /cgn1_9/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /cgn1_9/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /cgn1_9/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /cgn1_9/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /cgn1_9/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /cgn1_9/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /cgn1_9/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /cgn1_9/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /cgn1_9/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /cgn1_9/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /cgn1_9/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /cgn1_9/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /cgn1_9/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /cgn1_9/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /cgn1_9/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /cgn1_9/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /cgn1_9/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /cgn1_9/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	22	100.0	22	17	AAT27313	T.cruzi epimastigo
	2	22	100.0	22	20	AAX84095	PCR primer for T.
	3	22	100.0	3402	17	AAT27310	Trypanosoma cruzi
	4	22	100.0	3402	20	AAX84092	T. cruzi PTc40 cod
	5	16.2	73.6	123	21	AAC24763	Human secreted pro
c	6	16.2	73.6	10475	9	AAN80616	Genomic clone enco
c	7	16.2	73.6	10475	9	AAN80643	Sequence of human
	8	15.8	71.8	282	18	AAT91520	Mycobacterium tube
	9	15.8	71.8	282	18	AAT91454	Mycobacterium tube
	10	15.8	71.8	282	19	AAV64502	M. tuberculosis im
	11	15.8	71.8	282	19	AAV44394	Mycobacterium tube
	12	15.8	71.8	282	20	AAZ19304	M. tuberculosis an
	13	15.8	71.8	282	20	AAZ19092	M. tuberculosis re
c	14	15.8	71.8	404	22	AAF68508	Human lung tumour
	15	15.6	70.9	1661	21	AAA37040	Human PRO1772 (UNQ
	16	15.6	70.9	1661	22	AAF54245	DNA encoding prote
	17	15.6	70.9	2117	21	AAC76843	Human ORFX ORF2398
c	18	15.6	70.9	14327	22	AAF54724	Nucleotide sequenc
	19	15.4	70.0	1262	21	AAC39276	Arabidopsis thalia
	20	15.2	69.1	136	18	AAT49467	AAV sequence joine
	21	15.2	69.1	136	21	AAC09815	Human secreted pro
	22	15.2	69.1	355	18	AAT91514	Mycobacterium tube
	23	15.2	69.1	355	18	AAT91450	Mycobacterium tube
	24	15.2	69.1	355	19	AAV64497	M. tuberculosis im
	25	15.2	69.1	355	19	AAV44389	Mycobacterium tube
	26	15.2	69.1	355	20	AAZ19299	M. tuberculosis an
	27	15.2	69.1	355	20	AAZ19087	M. tuberculosis re
c	28	15.2	69.1	790	19	AAV24111	Human tumour prote
c	29	15.2	69.1	883	21	AAC56131	Eucalyptus grandis
c	30	15.2	69.1	1146	16	AAQ86990	Creatine-kinase su
c	31	15.2	69.1	1437	21	AAF18037	Lung cancer associ
	32	15.2	69.1	1539	19	AAV64478	M. tuberculosis im
	33	15.2	69.1	1539	19	AAV44370	Mycobacterium tube
	34	15.2	69.1	1539	20	AAZ19280	M. tuberculosis an
	35	15.2	69.1	1539	20	AAZ19068	M. tuberculosis re
c	36	15.2	69.1	1568	14	AAQ51620	Human creatine kin
c	37	15.2	69.1	1568	14	AAQ58990	Mutant human creat
	38	15.2	69.1	2064	12	AAQ14178	Gene encoding asco
c	39	15.2	69.1	2138	19	AAV64549	M. tuberculosis im
c	40	15.2	69.1	2138	19	AAV44440	Mycobacterium tube
c	41	15.2	69.1	2138	20	AAZ19350	M. tuberculosis an
c	42	15.2	69.1	2138	20	AAZ19138	M. tuberculosis re
c	43	15.2	69.1	2343	20	AAX99989	Creatine kinase CK
c	44	15.2	69.1	2438	21	AAF21626	Human breast and o
	45	15.2	69.1	4067	15	AAQ63192	Human adeno-associ

ALIGNMENTS

RESULT 1
AAT27313
ID AAT27313 standard; cDNA; 22 BP.
XX
AC AAT27313;
XX
DT 26-NOV-1996 (first entry)
XX
DE T.cruzi epimastigotic PTc100t antigen primer corresp. to bases 1266-87.
XX
KW Antigen; Trypanosoma cruzi; epimastigote; serum; Chagas disease; probe;
KW Primer; PCR; polymerase chain reaction; amplification; antibody; ss.
XX
OS Synthetic.
XX
PN FR2723589-A1.
XX
PD 16-FEB-1996.
XX
PF 12-AUG-1994; 94FR-0010132.
XX
PR 12-AUG-1994; 94FR-0010132.
XX
PA (INMR) BIO MERIEUX.
XX
PI Jolivet M, Lesenechal M, Paranhos-Baccala G;
XX
DR WPI; 1996-190287/20.
XX
PT New nucleic acid encoding Trypanosoma cruzi epimastigotic antigen -
PT useful for diagnosis, monitoring and therapy of Chagas disease
XX
PS Claim 29; Page 37; 55pp; French.
XX

```
CC   The primers AAT27311-5 were used to PCR amplify the sequence encoding a
CC   novel isolated antigenic protein from Trypanosoma cruzi epimastigotes,
CC   designated PTc100t (AAT27310). The primers, derived from the sequences
CC   of a 594 and 1041 bp fragment of PTc100t, amplified the gene as 3
CC   fragments. This primer corresponds to nucleotides 1266-87 of the PT100t
CC   sequence, derived from the 594 bp fragment. The 594 bp fragment was
CC   isolated from a T.cruzi genomic expression library in lambda gt11, using
CC   a mixture of sera from patients with Chagas disease. It corresponds to
CC   nucleotides 1232-1825 of PTc100t. The 1041 bp fragment was isolated from
CC   a lambda gt10 library using the 594 bp fragment as a probe. The protein,
CC   or antibodies raised against it, can be used in the detection and
CC   monitoring of T.cruzi infection i.e. Chagas disease.
XX
SQ   Sequence 22 BP; 3 A; 8 C; 7 G; 4 T; 0 other;


  Query Match            100.0%;  Score 22;  DB 17;  Length 22;
  Best Local Similarity  100.0%;  Pred. No. 0.12;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcgtcct 22
          ||||||||||||||||||||||
Db      1 cagccgacggtagctgcgtcct 22


RESULT   2
AAX84095
ID   AAX84095 standard; DNA; 22 BP.
XX
AC   AAX84095;
XX
DT   27-AUG-1999  (first entry)
XX
DE   PCR primer for T. cruzi PTc40 coding sequence.
XX
KW   PTc40; Tc40; infection; diagnosis; immune complex; antigenic determinant;
KW   therapy; antibody; PCR primer;  ss.
XX
OS   Synthetic.
OS   Trypanosoma cruzi.
XX
PN   WO9929867-A1.
XX
PD   17-JUN-1999.
XX
PF   10-DEC-1998;   98WO-IB01987.
XX
PR   10-DEC-1997;   97US-0988242.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Jolivet M,  Lesenechal M,  Mandrand B,  Paranhos-Baccala G;
XX
DR   WPI; 1999-394978/33.
XX
PT   New Trypanosoma cruzi antigen
XX
PS   Disclosure; Page 21; 65pp; English.
XX
CC   This sequence is a PCR primer for DNA encoding the Trypanosoma cruzi
CC   PTc40 protein of the invention. The PTc40 antigenic determinant is
CC   useful as a reagent for detection and/or monitoring of Trypanosoma cruzi
CC   infection from samples including blood serum or plasma, urine, saliva, or
CC   tears, by contacting with the sample and detecting an immune complex. The
CC   PTc40 antigenic determinant, the vector, expression cassette, cell or
CC   antibody are useful for treatment or prevention (vaccine) of a
CC   Trypanosoma cruzi infection in a man or animal. Current Trypanosoma cruzi
CC   antigens are obtained from protein fractions of the noninfectious stage
CC   of the parasite, and these do not allow sufficient production of antigens
CC   for use in reliable serological diagnostic tests. The strain to strain
CC   polymorphism reduces reliability of the tests.
XX
SQ   Sequence 22 BP; 3 A; 8 C; 7 G; 4 T; 0 other;


  Query Match            100.0%;  Score 22;  DB 20;  Length 22;
  Best Local Similarity  100.0%;  Pred. No. 0.12;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcgtcct 22
          ||||||||||||||||||||||
Db      1 cagccgacggtagctgcgtcct 22


RESULT   3
AAT27310
ID   AAT27310 standard; cDNA; 3402 BP.
XX
AC   AAT27310;
XX
DT   26-NOV-1996  (first entry)
XX
DE   Trypanosoma cruzi epimastigotic PTc100t antigen gene.
XX
KW   Antigen; Trypanosoma cruzi; epimastigote; serum; Chagas disease; probe;
KW   Primer; PCR; polymerase chain reaction; amplification; antibody; ds.
XX
OS   Trypanosoma cruzi.
XX
FH   Key             Location/Qualifiers
FT   CDS             266..3013
FT                   /*tag=  a
FT                   /product= PTc100t epimastogotic antigen
XX
PN   FR2723589-A1.
XX
PD   16-FEB-1996.
XX
PF   12-AUG-1994;   94FR-0010132.
XX
PR   12-AUG-1994;   94FR-0010132.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Jolivet M,  Lesenechal M,  Paranhos-Baccala G;
XX
DR   WPI; 1996-190287/20.
DR   P-PSDB; AAR91615.
XX
PT   New nucleic acid encoding Trypanosoma cruzi epimastigotic antigen -
PT   useful for diagnosis, monitoring and therapy of Chagas disease
XX
PS   Claim 1; Page 24-26; 55pp; French.
XX
CC   This is the nucleotide sequence encoding a novel isolated antigenic
CC   protein from Trypanosoma cruzi epimastigotes, designated PTc100t.
CC   The clone Tc50 was isolated from a T.cruzi genomic expression library in
CC   lambda gt11, using a mixture of sera from patients with Chagas disease.
CC   Clone Tc50 contained an 594 bp insert corresp. to nucleotides 1232-1825
CC   of this sequence.  The Tc50 sequence was subsequently used to probe a
CC   Southern blot of restriction enzyme digested T.cruzi DNA and also screen
CC   a lambda gt10 library to isolate a 1041 bp EcoRI fragment corresp. to
CC   nucleotides 1403-2443 of PTc100t.  Primers (AAT27311-5) were synthesised
CC   based on the sequences of the 594 and 1041 bp fragments and used to
CC   amplify the PTc100t clone as 3 fragments from cDNA derived from mRNA
CC   purified from T.cruzi epimastigotes.  The protein or antibodies raised
CC   against it can be used in the detection and monitoring of T.cruzi
CC   infection i.e. Chagas disease.
XX
SQ   Sequence 3402 BP; 889 A; 818 C; 958 G; 737 T; 0 other;


  Query Match            100.0%;  Score 22;  DB 17;  Length 3402;
  Best Local Similarity  100.0%;  Pred. No. 0.14;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      1 cagccgacggtagctgcgtcct 22
          ||||||||||||||||||||||
Db   1266 cagccgacggtagctgcgtcct 1287


RESULT   4
AAX84092
ID   AAX84092 standard; cDNA; 3402 BP.
XX
AC   AAX84092;
XX
DT   27-AUG-1999  (first entry)
XX
DE   T. cruzi PTc40 coding sequence.
XX
KW   PTc40; Tc40; infection; diagnosis; immune complex; antigenic determinant;
KW   therapy; antibody; ds.
XX
OS   Trypanosoma cruzi.
XX
PN   WO9929867-A1.
XX
PD   17-JUN-1999.
XX
PF   10-DEC-1998;   98WO-IB01987.
XX
PR   10-DEC-1997;   97US-0988242.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Jolivet M,  Lesenechal M,  Mandrand B,  Paranhos-Baccala G;
XX
DR   WPI; 1999-394978/33.
DR   P-PSDB; AAY22124.
XX
PT   New Trypanosoma cruzi antigen
XX
PS   Claim 1; Page 52-56; 65pp; English.
XX
CC   This sequence encodes the Trypanosoma cruzi PTc40 protein of the
CC   invention, and is designated Tc40. The PTc40 antigenic determinant is
CC   useful as a reagent for detection and/or monitoring of Trypanosoma cruzi
CC   infection from samples including blood serum or plasma, urine, saliva, or
CC   tears, by contacting with the sample and detecting an immune complex. The
CC   PTc40 antigenic determinant, the vector, expression cassette, cell or
CC   antibody are useful for treatment or prevention (vaccine) of a
CC   Trypanosoma cruzi infection in a man or animal. Current Trypanosoma cruzi
CC   antigens are obtained from protein fractions of the noninfectious stage
CC   of the parasite, and these do not allow sufficient production of antigens
CC   for use in reliable serological diagnostic tests. The strain to strain
CC   polymorphism reduces reliability of the tests.
XX
SQ   Sequence 3402 BP; 888 A; 821 C; 956 G; 737 T; 0 other;


  Query Match             100.0%;  Score 22;  DB 20;  Length 3402;
  Best Local Similarity   100.0%;  Pred. No. 0.14;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcgtcct 22
          ||||||||||||||||||||||
Db   1266 cagccgacggtagctgcgtcct 1287


RESULT   5
AAC24763
ID   AAC24763 standard; cDNA; 123 BP.
XX
AC   AAC24763;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 28838.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 28838; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 123 BP; 24 A; 44 C; 35 G; 20 T; 0 other;


  Query Match             73.6%;  Score 16.2;  DB 21;  Length 123;
  Best Local Similarity   85.7%;  Pred. No. 72;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcgtcc 21
          ||||||  || ||||||||||
Db     75 cagccggaggcagctgcgtcc 95


RESULT   6
AAN80616/c
ID   AAN80616 standard; DNA; 10475 BP.
XX
AC   AAN80616;
XX
DT   21-NOV-1990  (first entry)
XX
DE   Genomic clone encoding SAP(Phe).
XX
KW   SAP(Val); SAP(Phe); hyaline membrane disease; ss.
XX
FH   Key             Location/Qualifiers
FT   exon            1054..1120
FT                   /*tag=  a
FT                   /label=exon I
FT   exon            1431..1558
FT                   /*tag=  b
FT                   /label=exon II
FT   exon            2060..2131
```

```
FT                   /*tag= c
FT                   /label=exon III
FT   exon            2495..2620
FT                   /*tag= d
FT                   /label=exon IV
FT   exon            3447..3635
FT                   /*tag= e
FT                   /label=exon V
FT   exon            3863..3952
FT                   /*tag= f
FT                   /label=exon VI
FT   exon            5386..5568
FT                   /*tag= g
FT                   /label=exon VII
FT   exon            5742..5887
FT                   /*tag= h
FT                   /label=exon VIII
FT   exon            7122..7202
FT                   /*tag= i
FT                   /label=exon IX
FT   exon            7697..7756
FT                   /*tag= j
FT                   /label=exon X
XX
PN   WO8804324-A.
XX
PD   16-JUN-1988.
XX
PF   03-DEC-1987;   87WO-US03180.
XX
PR   03-DEC-1987;   87WO-US03180.
PR   03-DEC-1987;   87US-0939206.
XX
PA   (ABBO ) ABBOTT LABORATORIES.
XX
PI   Whitsett JA,  Fox JL,  Pilot-Matias TJ,  Meuth JL;
XX
DR   WPI; 1988-175472/25.
DR   P-PSDB; AAP82934.
XX
PT   Pulmonary hydrophobic surfactant-associated proteins - useful for
PT   normalising pulmonary surface tension.
XX
PS   Disclosure; ; pp; English.
XX
CC   This genomic clone encodes the pulmonary hydrophobic surfactant-
CC   associated protein SAP(Phe). Following processing to remove the
CC   introns the prod. is obtd. in a high yield.  When SAP(Phe) is
CC   combined with a lipid it can be used to reduce or maintain
CC   normal pulmonary surface tension in the alveoli of animals
CC   (esp. humans) and therefore can be used in the treatment of
CC   hyaline membrane disease in premature infants.
CC   May also be used to deliver substances to respiratory epithelial cells.
CC   See also AAP80572-96, AAN80613, AAN80615, AAN80617 and AAN82412-13.
XX
SQ   Sequence 10475 BP; 2416 A; 3063 C; 2940 G; 2056 T; 0 other;

  Query Match             73.6%;  Score 16.2;  DB 9;  Length 10475;
  Best Local Similarity   85.7%;  Pred. No. 83;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 agccgacggtagctgcgtcct 22
           |||| ||||| |||| |||||
Db    8369 AGCCCACGGTTGCTGTGTCCT 8349


RESULT   7
AAN80643/c
ID   AAN80643 standard; DNA; 10475 BP.
XX
AC   AAN80643;
XX
DT   03-FEB-1991  (first entry)
XX
DE   Sequence of human pulmonary hydrophobic surfactant-associated protein
DE   (SAP) (Phe) genomic clone.
XX
KW   Human pulmonary hydrophobic surfactant-associated protein (SAP) (Phe);
KW   hyaline membrane disease (HMD); therapy; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   exon            1054..1119
FT                   /*tag= a
FT                   /note="AAP80193"
FT   exon            1433..1558
FT                   /codon_start=1433
FT                   /*tag= b
FT                   /note="AAP80194"
FT   exon            2059..2130
FT                   /*tag= c
FT                   /note="AAN80195"
FT   exon            2494..2619
FT                   /*tag= d
FT                   /note="AAP81070"
FT   exon            3445..3633
FT                   /*tag= e
FT                   /note="AAP81074"
FT   exon            3861..3950
FT                   /*tag= f
FT                   /note="AAP81075"
FT   exon            5384..5567
FT                   /*tag= g
FT                   /note="AAP81076"
FT   exon            5742..5885
FT                   /*tag= h
FT                   /codon_start=5742
FT                   /note="AAP81077"
FT   exon            7120..7200
FT                   /*tag= i
FT                   /note="AAP81078"
FT   exon            7695..7754
FT                   /*tag= j
FT                   /note="AAP81079"
XX
PN   WO8803170-A.
XX
PD   05-MAY-1988.
XX
PF   02-OCT-1987;   87WO-US02536.
XX
PR   01-OCT-1987;   87US-0101680.
PR   08-DEC-1986;   86US-0939206.
PR   10-JUN-1987;   87US-0060719.
XX
PA   (WHIT/) WHITSETT JA.
XX
PI   Whitsett JA,  Fox JL,  Pilot-Matias TJ,  Meuth JL,  Sarin VK;
XX
DR   WPI; 1988-133244/19.
DR   P-PSDB; AAP80193, AAP80194, AAP80195, 081070, AAP81074, AAP81075,
DR   AAP81076, AAP81077, AAP81078, AAP81079.
XX
PT   Pulmonary hydrophobic surfactant-associated proteins -
PT   used with lipid(s) to treat and prevent hyaline membrane disease
PT   and similar syndromes
XX
PS   Example; Fig 8B-1 - 8B-24; 144pp; English.
XX
CC   The sequence is derived from human embryonic kidney cell genomic
CC   library. SAP (Val) and SAP (Phe), when combined with lipids, have
CC   significant pulmonary biophysical surfactant activity that may be
CC   utilised to treat and prevent hyaline membrane disease (HMD) and
CC   other syndromes associated with lack or insufficient amts. of natural
```

```
CC   pulmonary surfactant material.  Antibodies and antisera may also be
CC   made which are directed against SAP (Val) or SAP (Phe).
CC   SAP (Val) and SAP (Phe) in body fluids may be assayed using the compsns.
XX
SQ   Sequence 10475 BP; 2415 A; 3060 C; 2942 G; 2056 T; 2 other;


  Query Match           73.6%;  Score 16.2;  DB 9;  Length 10475;
  Best Local Similarity   85.7%;  Pred. No. 83;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 agccgacggtagctgcgtcct 22
          |||| ||||| |||| |||||
Db   8369 AGCCCACGGTTGCTGTGTCCT 8349


RESULT   8
AAT91520
ID   AAT91520 standard; DNA; 282 BP.
XX
AC   AAT91520;
XX
DT   09-JAN-1998  (first entry)
XX
DE   Mycobacterium tuberculosis antigen TbH-8-2 encoding DNA.
XX
KW   Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
KW   skin testing; M.tuberculosis; ss.
XX
OS   Mycobacterium tuberculosis.
XX
PN   WO9709428-A2.
XX
PD   13-MAR-1997.
XX
PF   30-AUG-1996;   96WO-US14674.
XX
PR   12-JUL-1996;   96US-0680574.
PR   01-SEP-1995;   95US-0523436.
PR   22-SEP-1995;   95US-0533634.
PR   22-MAR-1996;   96US-0620874.
PR   05-JUN-1996;   96US-0659683.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Campos-neto A,  Dillon DC,  Houghton R,  Reed SG,  Skeiky YAW;
PI   Twardzik DR, Vedvick TH;
XX
DR   WPI; 1997-192903/17.
XX
PT   New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are
PT   useful in vaccines for prevention or treatment of tuberculosis, also
PT   for diagnosis
XX
PS   Example 3; Page 136; 168pp; English.
XX
CC   A new immunogenic polypeptide has been developed comprising an
CC   immunogenic part of a soluble Mycobacterium tuberculosis antigen (or
CC   its variant differing only in conservative substitutions and/or
CC   modifications). The present sequence encodes a M.tuberculosis antigen,
CC   TbH-8-2. The immunogenic protein, and fusion proteins containing one
CC   or more of the proteins or one of the proteins plus ESAT-6, are useful
CC   in vaccines, preferably when formulated with a non-specific adjuvant,
CC   to induce an immune response against M.tuberculosis (for treatment or
CC   prevention).
XX
SQ   Sequence 282 BP; 56 A; 79 C; 86 G; 54 T; 7 other;


  Query Match           71.8%;  Score 15.8;  DB 18;  Length 282;
  Best Local Similarity   85.0%;  Pred. No. 1.1e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcgtc 20
          ||||||||||| ||| || |
Db    136 cagccgacggtggctccgnc 155


RESULT   9
AAT91454
ID   AAT91454 standard; DNA; 282 BP.
XX
AC   AAT91454;
XX
DT   13-JAN-1998  (first entry)
XX
DE   Mycobacterium tuberculosis antigen TbH-8-2 encoding DNA.
XX
KW   Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
KW   skin testing; M.tuberculosis; ss.
XX
OS   Mycobacterium tuberculosis.
XX
PN   WO9709429-A2.
XX
PD   13-MAR-1997.
XX
PF   30-AUG-1996;   96WO-US14675.
XX
PR   12-JUL-1996;   96US-0680573.
PR   01-SEP-1995;   95US-0523435.
PR   22-SEP-1995;   95US-0532136.
PR   22-MAR-1996;   96US-0620280.
PR   05-JUN-1996;   96US-0658800.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Campos-neto A,  Dillon DC,  Houghton R,  Reed SG,  Skeiky YAW;
PI   Twardzik DR, Vedvick TH;
XX
DR   WPI; 1997-192904/17.
XX
PT   New immunogenic polypeptide(s) from soluble M. tuberculosis antigens
PT   - useful for diagnosis of M. tuberculosis infection
XX
PS   Example 3; Page 148; 190pp; English.
XX
CC   A new immunogenic polypeptide has been developed comprising an
CC   immunogenic part of a soluble Mycobacterium tuberculosis antigen (or
CC   its variant differing only in conservative substitutions and/or
CC   modifications). The present sequence encodes a M.tuberculosis antigen,
CC   TbH-8-2. The immunogenic polypeptide can be used to diagnose
CC   M.tuberculosis infection by forming complexes with specific antibodies
CC   in the sample. Fragments of DNA encoding the immunogenic polypeptide
CC   can be used as diagnostic primers or probes and agents that bind to
CC   the antigen, especially monoclonal antibodies or equivalent polyclonal
CC   antibodies, are also used for diagnosis.
XX
SQ   Sequence 282 BP; 56 A; 79 C; 86 G; 54 T; 7 other;


  Query Match           71.8%;  Score 15.8;  DB 18;  Length 282;
  Best Local Similarity   85.0%;  Pred. No. 1.1e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcgtc 20
          ||||||||||| ||| || |
Db    136 cagccgacggtggctccgnc 155


RESULT  10
AAV64502
ID   AAV64502 standard; DNA; 282 BP.
XX
AC   AAV64502;
XX
```

DT 27-JAN-1999 (first entry)
XX
DE M. tuberculosis immunogenic polypeptide TbH-8-2 DNA.
XX
KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
KW vaccine; pharmaceutical; infection; diagnosis; ss.
XX
OS Mycobacterium tuberculosis.
XX
PN WO9816646-A2.
XX
PD 23-APR-1998.
XX
PF 07-OCT-1997; 97WO-US18293.
XX
PR 13-MAR-1997; 97US-0818112.
PR 11-OCT-1996; 96US-0730510.
XX
PA (CORI-) CORIXA CORP.
XX
PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
DR WPI; 1998-261042/23.
XX
PT Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used
PT to develop products for the detection of M. tuberculosis infection
PT and for diagnosis, treatment and prevention of tuberculosis
XX
PS Example 3b; Page 126; 230pp; English.
XX
CC This sequence encodes an immunogenic portion of a soluble Mycobacterium
CC tuberculosis (MT) antigen which can be used in a method for inducing
CC protective immunity against tuberculosis (TB). This sequence can be
CC formulated into vaccines and/or pharmaceutical compositions for
CC immunising against M. tuberculosis infection or may be used for the
CC diagnosis of tuberculosis.
XX
SQ Sequence 282 BP; 56 A; 79 C; 86 G; 54 T; 7 other;

Query Match 71.8%; Score 15.8; DB 19; Length 282;
Best Local Similarity 85.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 cagccgacggtagctgcgtc 20
||||||||||| ||| || |
Db 136 cagccgacggtggctccgnc 155

RESULT 11
AAV44394
ID AAV44394 standard; DNA; 282 BP.
XX
AC AAV44394;
XX
DT 09-NOV-1998 (first entry)
XX
DE Mycobacterium tuberculosis antigen TbH-8-2 DNA.
XX
KW Tuberculosis; infection; diagnosis; antigen; TbH-8-2; ss.
XX
OS Mycobacterium tuberculosis strain H37Rv.
XX
PN WO9816645-A2.
XX
PD 23-APR-1998.
XX
PF 07-OCT-1997; 97WO-US18214.
XX
PR 13-MAR-1997; 97US-0818111.
PR 11-OCT-1996; 96US-0729622.
XX
PA (CORI-) CORIXA CORP.
XX
PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
DR WPI; 1998-251292/22.
XX
PT New isolated Mycobacterium tuberculosis polypeptides and DNA - used
PT to develop products for the detection of M. tuberculosis infection
PT and diagnosis of tuberculosis
XX
PS Example 3; Page 131; 250pp; English.
XX
CC This DNA sequence codes for an antigenic portion of Mycobacterium
CC tuberculosis antigen TbH-8-2. It was isolated from a M.
CC tuberculosis strain H37Rv expression library using sera from
CC patients having pulmonary or pleural tuberculosis, and is a partial
CC clone of TbH-8 (see AAV44370). The invention relates to compositions
CC and methods for diagnosing tuberculosis. It provides polypeptides
CC (see AAW64291-W64379) comprising an antigenic portion of a soluble M.
CC tuberculosis antigen, or an immunogenic portion of an M.
CC tuberculosis antigen, as well as DNA sequences encoding such
CC polypeptides, recombinant expression vectors and transformed or
CC transfected host cells. Also claimed are methods and diagnostic
CC kits for detecting M. tuberculosis infection in a patient using
CC these polypeptides, antibodies or oligonucleotide probes and
CC primers, for the diagnosis of tuberculosis.
XX
SQ Sequence 282 BP; 56 A; 79 C; 86 G; 54 T; 7 other;

Query Match 71.8%; Score 15.8; DB 19; Length 282;
Best Local Similarity 85.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 cagccgacggtagctgcgtc 20
||||||||||| ||| || |
Db 136 cagccgacggtggctccgnc 155

RESULT 12
AAZ19304
ID AAZ19304 standard; DNA; 282 BP.
XX
AC AAZ19304;
XX
DT 05-NOV-1999 (first entry)
XX
DE M. tuberculosis antigen TbH-8-2 DNA sequence.
XX
KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
KW immunotherapy; diagnosis; immunisation; vaccine; infection;
KW immune response; skin test; ss.
XX
OS Mycobacterium tuberculosis.
XX
PN WO9942076-A2.
XX
PD 26-AUG-1999.
XX
PF 17-FEB-1999; 99WO-US03268.
XX
PR 05-MAY-1998; 98US-0072967.
PR 18-FEB-1998; 98US-0025197.
XX
PA (CORI-) CORIXA CORP.
XX
PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
DR WPI; 1999-527409/44.
XX
PT New antigens from Mycobacterium tuberculosis useful in diagnostic

```
PT   skin tests and protective or therapeutic vaccines or compositions
XX
PS   Example 3; Page 120; 299pp; English.
XX
CC   The present invention describes polypeptides comprising an immunogenic
CC   part of a Mycobacterium tuberculosis antigen (Ag). Also described
CC   are vaccines and fusion protein containing M. tuberculosis Ag's.
CC   M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
CC   other polypeptides fragments, can be used in pharmaceutical compositions
CC   or vaccines to generate a protective or therapeutic immune response to
CC   M. tuberculosis and as reagents in skin tests for diagnosis of
CC   tuberculosis. Ag can induce proliferation of, or cytokine secretion
CC   by, T, B or natural killer cells and/or macrophages in
CC   tuberculosis-immune subjects. AAZ19249 to AAZ19460 and AAY39083 to
CC   AAY39225 are used in the exemplification of the present invention.
XX
SQ   Sequence 282 BP; 56 A; 79 C; 86 G; 54 T; 7 other;


  Query Match             71.8%;  Score 15.8;  DB 20;  Length 282;
  Best Local Similarity   85.0%;  Pred. No. 1.1e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcgtc 20
          ||||||||||| ||| || |
Db    136 cagccgacggtggctccgnc 155


RESULT  13
AAZ19092
ID   AAZ19092 standard; DNA; 282 BP.
XX
AC   AAZ19092;
XX
DT   05-NOV-1999  (first entry)
XX
DE   M. tuberculosis recombinant antigen DNA encoding TbH-8-2.
XX
KW   Antigen; diagnosis; detection; infection; antibody; immunisation;
KW   vaccine; immunity; ss.
XX
OS   Mycobacterium tuberculosis.
XX
PN   WO9942118-A2.
XX
PD   26-AUG-1999.
XX
PF   17-FEB-1999;   99WO-US03265.
XX
PR   05-MAY-1998;   98US-0072596.
PR   18-FEB-1998;   98US-0024753.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Campos-Neto A,  Dillon DC,  Hendrickson RC,  Houghton R;
PI   Lodes MJ,  Reed SG,  Skeiky YAW,  Twardzik DR,  Vedvick TS;
XX
DR   WPI; 1999-527416/44.
XX
PT   New polypeptide comprising antigenic portions of M. tuberculosis
XX
PS   Example 3; Page 165-166; 323pp; English.
XX
CC   This invention describes novel recombinant antigens and their encoding
CC   nucleic acids derived from Mycobacterium tuberculosis. The novel
CC   polypeptides are useful for detecting M. tuberculosis infection in a
CC   biological sample by detecting antibodies which bind with the
CC   polypeptides, and are useful as vaccines for immunizing against
CC   M. tuberculosis infection. The new detection methods are needed as
CC   current vaccination strategies do not provide 100% immunity.
XX
SQ   Sequence 282 BP; 56 A; 79 C; 86 G; 54 T; 7 other;


  Query Match             71.8%;  Score 15.8;  DB 20;  Length 282;
  Best Local Similarity   85.0%;  Pred. No. 1.1e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcgtc 20
          ||||||||||| ||| || |
Db    136 cagccgacggtggctccgnc 155


RESULT  14
AAF68508/c
ID   AAF68508 standard; cDNA; 404 BP.
XX
AC   AAF68508;
XX
DT   12-APR-2001  (first entry)
XX
DE   Human lung tumour protein related nucleotide sequence SEQ ID NO:443.
XX
KW   Human; lung cancer; lung tumour; lung tumour protein; gene therapy;
KW   lung cancer antigen; lung tumour-specific antigen; diagnosis; vaccine;
KW   cytostatic; antisense inhibition; ss.
XX
OS   Homo sapiens.
XX
PN   WO200100828-A2.
XX
PD   04-JAN-2001.
XX
PF   30-JUN-2000; 2000WO-US18061.
XX
PR   30-JUN-1999;   99US-0346492.
PR   15-OCT-1999;   99US-0419356.
PR   17-DEC-1999;   99US-0466867.
PR   30-DEC-1999;   99US-0476300.
PR   06-MAR-2000; 2000US-0519642.
PR   22-MAR-2000; 2000US-0533077.
PR   10-APR-2000; 2000US-0546259.
PR   27-APR-2000; 2000US-0560406.
PR   05-JUN-2000; 2000US-0589184.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Wang T,  Bangur CS,  Lodes MJ,  Fanger GR,  Vedvick TS,  Carter D;
PI   Retter MW,  Mannion J;
XX
DR   WPI; 2001-071488/08.
XX
PT   Lung tumor-associated proteins and the nucleic acids that encode them,
PT   useful for preventing, diagnosing and treating lung cancer -
XX
PS   Claim 4; Page 302; 436pp; English.
XX
CC   The present invention describes immunogenic portions of lung tumour-
CC   associated proteins (I) and the nucleic acids (NAs) that encode them.
CC   (I) have cytostatic activity and can be used in gene therapy, antisense
CC   inhibition and in vaccines. The NAs and the lung tumour-associated
CC   proteins they encode may be used in the prevention, treatment and
CC   diagnosis of diseases associated with their inappropriate expression,
CC   especially lung cancers. For example, the NAs may be administered to
CC   treat diseases by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of the protein by expressing inactive proteins
CC   or to supplement the patients own production of (I). Additionally, the
CC   NAs may be used to produce the lung-tumour associated protein, according
CC   to standard recombinant DNA methodology. Conversely, antisense NA
CC   molecules may be administered to down regulate protein expression by
CC   binding with the cells own genes and preventing their expression. The NA
CC   and complementary sequences may also be used as DNA probes in diagnostic
CC   assays to detect and quantitate the presence of similar NA sequences in
CC   samples, and hence which patients may be in need of treatment for lung
CC   cancer. The (I) may be used as antigens in the production of antibodies
CC   and in assays to identify modulators (agonists and antagonists) of the
```

```
CC   expression and activity of the protein. AAF68083 to AAF68878 and
CC   AAB76848 to AAB76878 represent human lung tumour protein related
CC   nucleotide and protein sequences which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 404 BP; 87 A; 112 C; 112 G; 85 T; 8 other;


  Query Match            71.8%;  Score 15.8;  DB 22;  Length 404;
  Best Local Similarity  81.0%;  Pred. No. 1.2e+02;
  Matches   17;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 agccgacggtagctgcgtcct 22
           ||||  |||| |||| |||||
Db     316 AGCCCNCGGTNGCTGTGTCCT 296


RESULT  15
AAA37040
ID   AAA37040 standard; cDNA; 1661 BP.
XX
AC   AAA37040;
XX
DT   08-AUG-2000  (first entry)
XX
DE   Human PRO1772 (UNQ834) cDNA sequence SEQ ID NO:62.
XX
KW   Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
KW   transmembrane; secretion; immunoadhesion; pharmaceutical; screening;
KW   ss.
XX
OS   Homo sapiens.
XX
PN   WO200012708-A2.
XX
PD   09-MAR-2000.
XX
PF   01-SEP-1999;   99WO-US20111.
XX
PR   01-SEP-1998;   98US-0098716.
PR   01-SEP-1998;   98US-0098749.
PR   01-SEP-1998;   98US-0098750.
PR   02-SEP-1998;   98US-0098803.
PR   02-SEP-1998;   98US-0098821.
PR   02-SEP-1998;   98US-0098843.
PR   09-SEP-1998;   98US-0099536.
PR   09-SEP-1998;   98US-0099596.
PR   09-SEP-1998;   98US-0099598.
PR   09-SEP-1998;   98US-0099602.
PR   09-SEP-1998;   98US-0099642.
PR   10-SEP-1998;   98US-0099741.
PR   10-SEP-1998;   98US-0099754.
PR   10-SEP-1998;   98US-0099763.
PR   10-SEP-1998;   98US-0099792.
PR   10-SEP-1998;   98US-0099808.
PR   10-SEP-1998;   98US-0099812.
PR   10-SEP-1998;   98US-0099815.
PR   10-SEP-1998;   98US-0099816.
PR   15-SEP-1998;   98US-0100385.
PR   15-SEP-1998;   98US-0100388.
PR   15-SEP-1998;   98US-0100390.
PR   16-SEP-1998;   98US-0100584.
PR   16-SEP-1998;   98US-0100627.
PR   16-SEP-1998;   98US-0100661.
PR   16-SEP-1998;   98US-0100662.
PR   16-SEP-1998;   98US-0100664.
PR   17-SEP-1998;   98US-0100683.
PR   17-SEP-1998;   98US-0100684.
PR   17-SEP-1998;   98US-0100710.
PR   17-SEP-1998;   98US-0100711.
PR   17-SEP-1998;   98US-0100919.
PR   17-SEP-1998;   98US-0100930.
PR   18-SEP-1998;   98US-0100848.
PR   18-SEP-1998;   98US-0100849.
PR   18-SEP-1998;   98US-0101014.
PR   18-SEP-1998;   98US-0101068.
PR   18-SEP-1998;   98US-0101071.
PR   22-SEP-1998;   98US-0101279.
PR   23-SEP-1998;   98US-0101471.
PR   23-SEP-1998;   98US-0101472.
PR   23-SEP-1998;   98US-0101474.
PR   23-SEP-1998;   98US-0101475.
PR   23-SEP-1998;   98US-0101476.
PR   23-SEP-1998;   98US-0101477.
PR   23-SEP-1998;   98US-0101479.
PR   24-SEP-1998;   98US-0101738.
PR   24-SEP-1998;   98US-0101741.
PR   24-SEP-1998;   98US-0101743.
PR   24-SEP-1998;   98US-0101915.
PR   24-SEP-1998;   98US-0101916.
PR   29-SEP-1998;   98US-0102207.
PR   29-SEP-1998;   98US-0102240.
PR   29-SEP-1998;   98US-0102307.
PR   29-SEP-1998;   98US-0102330.
PR   29-SEP-1998;   98US-0102331.
PR   30-SEP-1998;   98US-0102484.
PR   30-SEP-1998;   98US-0102487.
PR   30-SEP-1998;   98US-0102570.
PR   30-SEP-1998;   98US-0102571.
PR   01-OCT-1998;   98US-0102684.
PR   01-OCT-1998;   98US-0102687.
PR   02-OCT-1998;   98US-0102965.
PR   06-OCT-1998;   98US-0103258.
PR   06-OCT-1998;   98US-0103449.
PR   07-OCT-1998;   98US-0103314.
PR   07-OCT-1998;   98US-0103315.
PR   07-OCT-1998;   98US-0103328.
PR   07-OCT-1998;   98US-0103395.
PR   07-OCT-1998;   98US-0103396.
PR   07-OCT-1998;   98US-0103401.
PR   08-OCT-1998;   98US-0103633.
PR   08-OCT-1998;   98US-0103678.
PR   08-OCT-1998;   98US-0103679.
PR   08-OCT-1998;   98US-0103711.
PR   14-OCT-1998;   98US-0104257.
PR   20-OCT-1998;   98US-0104987.
PR   20-OCT-1998;   98US-0105000.
PR   20-OCT-1998;   98US-0105002.
PR   21-OCT-1998;   98US-0105104.
PR   22-OCT-1998;   98US-0105169.
PR   22-OCT-1998;   98US-0105266.
PR   26-OCT-1998;   98US-0105693.
PR   26-OCT-1998;   98US-0105694.
PR   27-OCT-1998;   98US-0105807.
PR   27-OCT-1998;   98US-0105881.
PR   27-OCT-1998;   98US-0105882.
PR   27-OCT-1998;   98US-0106062.
PR   28-OCT-1998;   98US-0106023.
PR   28-OCT-1998;   98US-0106029.
PR   28-OCT-1998;   98US-0106030.
PR   28-OCT-1998;   98US-0106032.
PR   28-OCT-1998;   98US-0106033.
PR   28-OCT-1998;   98US-0106178.
PR   29-OCT-1998;   98US-0106248.
PR   29-OCT-1998;   98US-0106384.
PR   29-OCT-1998;   98US-0108500.
PR   30-OCT-1998;   98US-0106464.
PR   03-NOV-1998;   98US-0106856.
PR   03-NOV-1998;   98US-0106902.
PR   03-NOV-1998;   98US-0106905.
PR   03-NOV-1998;   98US-0106919.
PR   03-NOV-1998;   98US-0106932.
PR   03-NOV-1998;   98US-0106934.
PR   10-NOV-1998;   98US-0107783.
PR   17-NOV-1998;   98US-0108775.
PR   17-NOV-1998;   98US-0108779.
```

PR 17-NOV-1998; 98US-0108787.
PR 17-NOV-1998; 98US-0108788.
PR 17-NOV-1998; 98US-0108801.
PR 17-NOV-1998; 98US-0108802.
PR 17-NOV-1998; 98US-0108806.
PR 17-NOV-1998; 98US-0108807.
PR 17-NOV-1998; 98US-0108867.
PR 17-NOV-1998; 98US-0108925.
PR 18-NOV-1998; 98US-0108848.
PR 18-NOV-1998; 98US-0108849.
PR 18-NOV-1998; 98US-0108850.
PR 18-NOV-1998; 98US-0108851.
PR 18-NOV-1998; 98US-0108852.
PR 18-NOV-1998; 98US-0108858.
PR 18-NOV-1998; 98US-0108904.
XX
PA (GETH) GENENTECH INC.
XX
PI Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;
XX
DR WPI; 2000-237871/20.
DR P-PSDB; AAY99358.
XX
PT New mammalian DNA sequences encoding transmembrane, receptor or
PT secreted PRO polypeptides, useful for screening of potential peptide or
PT small molecule inhibitors of the relevant receptor/ligand interactions
XX
PS Claim 2; Fig 37; 773pp; English.
XX
CC AAA37022 to AAA37144 encode the new isolated human transmembrane,
CC receptor or secreted PRO polypeptides given in AAY99340 to AAY99462. The
CC transmembrane and receptor PRO proteins can be used for screening of
CC potential peptide or small molecule inhibitors of the relevant
CC receptor/ligand interactions. The polypeptides and nucleotide sequences
CC encoding then have various industrial applications, including uses as
CC pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent
CC PCR primers and hybridisation probes used in the isolation of the PRO
CC polypeptides from the present invention.
XX
SQ Sequence 1661 BP; 334 A; 519 C; 457 G; 349 T; 2 other;

Query Match 70.9%; Score 15.6; DB 21; Length 1661;
Best Local Similarity 81.8%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy      1 cagccgacggtagctgcgtcct 22
          |||||| ||||| ||||| | |
Db    128 cagccggcggtatctgcggcgt 149
```

Search completed: September 21, 2001, 02:42:41
Job time: 11457 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: September 20, 2001, 23:28:03 ; Search time 4309.39 Seconds
(without alignments)
48.258 Million cell updates/sec

Title: US-09-138-735-9
Perfect score: 22
Sequence: 1 cagccgacggtagctgcgtcct 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est75:*
137: gb_est76:*
138: gb_est77:*
139: gb_est78:*
140: gb_est79:*
141: gb_est80:*
142: gb_est81:*
143: gb_est82:*
144: gb_est83:*
145: gb_est84:*
146: gb_est85:*
147: gb_est86:*
148: gb_est87:*
149: gb_est88:*
150: gb_est89:*
151: gb_est90:*
152: gb_est99:*
153: gb_est100:*
154: gb_est101:*
155: gb_est102:*
156: gb_est103:*
157: gb_est104:*
158: gb_est105:*
159: gb_est106:*
160: gb_est67:*
161: gb_est68:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_esthum29:*
177: em_esthum30:*
178: em_esthum31:*
179: em_esthum32:*
180: em_esthum33:*
181: em_estom3:*
182: em_estpl11:*
183: em_estro21:*
184: em_estro22:*
185: em_estro23:*
186: em_htc:*
187: gb_est107:*
188: gb_est108:*
189: gb_est109:*
190: gb_est110:*
191: gb_est111:*
192: gb_htc:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_other:*
207: em_gss_pln1:*
208: em_gss_pln2:*
209: em_gss_pro:*
210: em_gss_rod1:*
211: em_gss_rod2:*
212: em_gss_rod3:*
213: em_gss_rod4:*
214: em_gss_rod5:*
215: em_gss_vrt1:*
216: em_gss_vrt2:*
217: em_gss_vrt3:*
218: gb_gss1:*
219: gb_gss2:*
220: gb_gss3:*
221: gb_gss4:*
222: gb_gss5:*
223: gb_gss6:*
224: gb_gss7:*
225: gb_gss8:*
226: gb_gss9:*
227: gb_gss10:*
228: gb_gss11:*
229: gb_gss12:*
230: gb_gss13:*
231: gb_gss14:*
232: gb_gss15:*
233: gb_gss16:*
234: gb_gss17:*
235: gb_gss18:*
236: gb_gss19:*
237: gb_gss20:*
238: gb_gss21:*
239: gb_gss22:*
240: gb_gss23:*
241: gb_gss24:*
242: gb_gss25:*
243: gb_gss26:*
244: gb_gss27:*
245: gb_gss28:*
246: gb_gss29:*
247: gb_gss30:*
248: gb_gss31:*
249: gb_gss32:*
250: gb_gss33:*
251: gb_gss34:*
252: em_gss_inv4:*
253: em_gss_rod6:*
254: em_gss_rod7:*
255: em_gss_rod8:*
256: gb_gss35:*
257: gb_gss36:*
258: gb_gss37:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	17.4	79.1	461	175	BG270722	BG270722 ia97h10.y
2	17.2	78.2	177	21	AI496394	AI496394 sb04b07.y
3	17.2	78.2	205	114	AW306389	AW306389 se50a06.y
4	17.2	78.2	452	234	AQ872227	AQ872227 nbeb00470
5	17.2	78.2	481	152	BG365834	BG365834 HVSMEi000
6	17.2	78.2	540	19	AI392116	AI392116 NCC2B4T3
7	17.2	78.2	594	121	AW861895	AW861895 CM3-CT034
8	17.2	78.2	744	14	AA950463	AA950463 LD30267.5
9	17.2	78.2	910	153	BG414942	BG414942 HVSMEk000
c 10	17.2	78.2	958	142	BE962535	BE962535 601655848
c 11	17.2	78.2	993	222	CNS05ORJ	AL347032 Tetraodon
12	16.8	76.4	259	113	AW265763	AW265763 xp81e09.x
13	16.8	76.4	281	170	BF825637	BF825637 RC0-HN002
c 14	16.8	76.4	332	107	AU032424	AU032424 AU032424
c 15	16.4	74.5	928	172	BF978349	BF978349 602148933
c 16	16.2	73.6	209	122	AW922321	AW922321 DG1_17_E0
c 17	16.2	73.6	246	116	AW445732	AW445732 82893 MAR
c 18	16.2	73.6	295	116	AW478058	AW478058 18366 MAR
19	16.2	73.6	338	222	FR0031088	AL027457 Fugu rubr
c 20	16.2	73.6	357	224	AQ073528	AQ073528 EP(2)2366
c 21	16.2	73.6	366	116	AW435588	AW435588 74113 MAR
c 22	16.2	73.6	385	115	AW418466	AW418466 56377 MAR
c 23	16.2	73.6	395	158	H48252	H48252 yq69a01.s1
c 24	16.2	73.6	402	116	AW463080	AW463080 BP230011A
c 25	16.2	73.6	403	116	AW447181	AW447181 88385 MAR
c 26	16.2	73.6	428	143	BF044716	BF044716 BP250001B
c 27	16.2	73.6	451	139	BE758397	BE758397 SWYACAL10
c 28	16.2	73.6	455	162	BE050606	BE050606 za67g01.b
c 29	16.2	73.6	457	16	AI105748	AI105748 SWAMCAC25
c 30	16.2	73.6	458	136	BE499187	BE499187 WHE0970_C
c 31	16.2	73.6	460	162	BE050607	BE050607 za67g01.g
32	16.2	73.6	464	117	AW510932	AW510932 hd41d12.x
33	16.2	73.6	471	22	AI566236	AI566236 tq70b06.x
c 34	16.2	73.6	472	143	BF043469	BF043469 BP250025B
c 35	16.2	73.6	477	147	BF358274	BF358274 PM1-EN000
c 36	16.2	73.6	485	188	R61321	R61321 yh07e07.s1
c 37	16.2	73.6	490	114	AW309353	AW309353 sf16c09.x
c 38	16.2	73.6	497	171	BF909469	BF909469 IL2-UT007
c 39	16.2	73.6	520	143	BF042379	BF042379 BP250022A
40	16.2	73.6	521	4	AA240489	AA240489 mv32b07.r
41	16.2	73.6	521	220	CNS02LDP	AL202678 Tetraodon
c 42	16.2	73.6	526	139	BE757430	BE757430 211865 MA
c 43	16.2	73.6	533	104	AI943345	AI943345 fc79a04.y
44	16.2	73.6	538	19	AI403626	AI403626 GH23107.5
c 45	16.2	73.6	540	119	AW657127	AW657127 109776 MA

ALIGNMENTS

RESULT 1
BG270722/c
LOCUS BG270722 461 bp mRNA EST 20-FEB-2001
DEFINITION ia97h10.y1 Melton Mouse Adult Pancreas 1 Mus musculus cDNA 5' similar to SW:TRY3_RAT P08426 TRYPSINOGEN III, CATIONIC PRECURSOR ;, mRNA sequence.
ACCESSION BG270722
VERSION BG270722.1 GI:12978157
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 461)
AUTHORS Melton,D., Meadows,A., Clifton,S., Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A., Schmitt,A., Theising,B., Ritter ,E., Ronko,I., Bennett,J., Cardenas,M., Gibbons,M., McCann,R., Cole ,R., Tsagareishvili,R., Williams,T., Jackson,Y. and Bowers,Y.
TITLE WashU-Harvard Pancreas EST Project
JOURNAL Unpublished (2000)
COMMENT Contact: Douglas Melton
WashU-Harvard Pancreas EST Project
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu, meadows@fas.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Juliana Brown (brown@fas.harvard.edu)
Seq primer: -40RP from Gibco
High quality sequence stop: 445.

FEATURES Location/Qualifiers
source 1. .461
/organism="Mus musculus"
/strain="ICR"
/db_xref="taxon:10090"
/clone_lib="Melton Mouse Adult Pancreas 1"
/sex="Male"
/tissue_type="Total Pancreas"
/dev_stage="Adult".
/lab_host="TOP10"
/note="Organ: Pancreas; Vector: pZErO-2; Site_1: Not I; Site_2: Xho I; Library constructed using SuperScript Plasmid Library kit (Life Technologies). cDNA made by oligo-dT priming. Xho I site destroyed during cloning. Size-selected by column fractionation. Primary library, unamplified."

BASE COUNT 92 a 118 c 158 g 84 t 9 others
ORIGIN

Query Match 79.1%; Score 17.4; DB 175; Length 461;
Best Local Similarity 85.7%; Pred. No. 3e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy      1 cagccgacggtagctgcgtcc 21
          ||||||||  ||||||  ||||
Db    202 CAGCCGACTGTAGCTNNGTCC 182
```

RESULT 2
AI496394
LOCUS AI496394 177 bp mRNA EST 01-DEC-1999
DEFINITION sb04b07.y1 Gm-c1004 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1004-7790 5', mRNA sequence.
ACCESSION AI496394
VERSION AI496394.1 GI:4397397
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE 1 (bases 1 to 177)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann ,R., Waterston,R. and Wilson,R.
TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800

```
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134 For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
            427-3324 or contact: clones@genomesystems.com or
            info@genomesystems.com web site: www.genomesystems.com
            Seq primer: -40RP from Gibco.
FEATURES            Location/Qualifiers
     source         1. .177
                    /organism="Glycine max"
                    /db_xref="taxon:3847"
                    /clone="GENOME SYSTEMS CLONE ID: Gm-c1004-7790"
                    /clone_lib="Gm-c1004"
                    /tissue_type="root"
                    /lab_host="XL10-Gold"
                    /note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
                    XhoI; Root cDNA. The mRNA was isolated from entire roots
                    of 8 day old 'Williams' seedlings which were propagated on
                    paper towels with distilled water. Stratagene's cDNA
                    Synthesis Kit (catalog #200401) was used to synthesize the
                    cDNA. First- strand synthesis was performed with 5-methyl
                    dCTP, hence the ligated cDNA is hemimethylated.
                    Stratagene's first-strand synthesis primer was used
                    [GAGAGAGAGAGAGAGAGAACTAGTCTCGAG(T)-18]. After
                    second-strand synthesis, the cDNA ends were 'polished'
                    with clone Pfu DNA polymerase, ligated to EcoRI adapters,
                    and phosphorylated. The XhoI site within the first-strand
                    synthesis primer was restricted by digestion with XhoI;
                    all XhoI sites in the cDNA would be protected by their
                    hemimethylated status. The cDNA constructs were
                    size-fractionated with a 500bp cutoff, using GibcoBRL Life
                    Technologies' cDNA Size Fractionation column. The column
                    eluent was then ligated into Stratagene's pBluescript II
                    XR Predigested vector (pBluescript II SK(+) that had been
                    digested with EcoRI and XhoI, and phosphorylated). Both
                    the white and blue colonies appear to contain recombinant
                    plasmids with cDNA inserts. Blue colonies 9n=15) have been
                    sequenced, and possess putative cDNA inserts. This library
                    was constructed by Dr. Paul Keim, & Virginia H. Coryell,
                    Department of Biology; Box5640, Northern Arizona
                    University, Flagstaff, AZ 86011, Phone: 520-523-1078 (Dr.
                    Paul Keim), 520-523-1372 (Virginia H. Coryell), Fax:
                    520-523-7500, email: paul.keim@nau.edu,
                    virginia.coryell@nau.edu"
BASE COUNT       28 a     62 c     25 g     62 t
ORIGIN


  Query Match             78.2%;  Score 17.2;  DB 21;  Length 177;
  Best Local Similarity   86.4%;  Pred. No. 3.8e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 cagccgacggtagctgcgtcct 22
           ||||||||||| ||||| | ||
Db     150 CAGCCGACGGTGGCTGCCTTCT 171


RESULT   3
AW306389
LOCUS       AW306389      205 bp    mRNA            EST       20-JAN-2000
DEFINITION  se50a06.y1 Gm-c1017 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1017-2315 5', mRNA sequence.
ACCESSION   AW306389
VERSION     AW306389.1  GI:6718742
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE   1  (bases 1 to 205)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134 For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
            427-3324 or contact: clones@genomesystems.com or
            info@genomesystems.com web site: www.genomesystems.com
            Seq primer: -40RP from Gibco.
FEATURES            Location/Qualifiers
     source         1. .205
                    /organism="Glycine max"
                    /db_xref="taxon:3847"
                    /clone="GENOME SYSTEMS CLONE ID: Gm-c1017-2315"
                    /clone_lib="Gm-c1017"
                    /tissue_type="vegetable buds of field grown plants"
                    /lab_host="XL10-Gold"
                    /note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
                    XhoI; This cDNA library was constructed from mRNA isolated
                    from vegatable buds of field grown plants. The cDNA
                    library was prepared using the Stratagene pBluescript II
                    XR library construction kit. Complementary DNA was
                    synthesized from mRNA using a primer consisting of a poly
                    (dT) sequence with a XhoI restriction site. EcoRI adapters
                    were ligated to the blunt-ended cDNA fragments followed by
                    XhoI digestion. The cDNA fragments were directionally
                    cloned into the EcoRI-XhoI restriction site of the
                    pBluescript vector. The ligated cDNA fragments were
                    transformed into XL10-Gold host cells. This library was
                    constructed by Dr. Randy Shoemaker and Dr. John
                    Erpelding."
BASE COUNT       59 a     47 c     41 g     58 t
ORIGIN


  Query Match             78.2%;  Score 17.2;  DB 114;  Length 205;
  Best Local Similarity   86.4%;  Pred. No. 3.8e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 cagccgacggtagctgcgtcct 22
           ||||||||||| ||||| | ||
Db      31 CAGCCGACGGTGGCTGCCTTCT 52


RESULT   4
AQ872227
LOCUS       AQ872227      452 bp    DNA             GSS       03-NOV-1999
DEFINITION  nbeb0047019r CUGI Rice BAC Library (EcoRI) Oryza sativa genomic
            clone nbeb0047019r, DNA sequence.
ACCESSION   AQ872227
VERSION     AQ872227.1  GI:6222678
KEYWORDS    GSS.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 452)
  AUTHORS   Wing,R.A. and Dean,R.A.
  TITLE     A BAC End Sequencing Framework to Sequence the Rice Genome
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Wing RA
```

```
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: GGAAACAGCTATGACCATG
            Class: BAC ends
            High quality sequence start: 32
            High quality sequence stop: 395.
FEATURES              Location/Qualifiers
     source           1. .452
                      /organism="Oryza sativa"
                      /strain="Japonica"
                      /cultivar="Nipponbare"
                      /db_xref="taxon:4530"
                      /clone="nbeb0047O19r"
                      /clone_lib="CUGI Rice BAC Library (EcoRI)"
                      /tissue_type="Leaf"
                      /lab_host="E. coli DH10B"
                      /note="Vector: pBACIndigo; Site_1: EcoRI; Site_2: EcoRI;
                      Rice is the most important food crop in the world. Half of
                      the world population, especially those inhabiting highly
                      populated areas of the humid tropics and subtropics, rely
                      on rice as their primary source of carbohydrate.
                      Monocotyledonous rice is a diploid plant (2n=24) with a
                      haploid genome equivalent of 431 Mbp (Arumuganathan and
                      Earle, 1991). The relatively small genome of rice, three
                      times larger than that of Arabidopsis, makes it suitable
                      for genomic studies. In order to facilitate positional
                      cloning, physical mapping and genome sequencing of rice,
                      we have constructed a BAC library from Oryza sativa,
                      Nipponbare variety using EcoRI as the cloning enzyme. The
                      library contains 55,296 clones with an average insert size
                      of 121 Kb providing approximatley 15 haploid genome
                      equivalents. The deep coverage allows the isolation a
                      particular sequence with a probability of 99.9 %. Three
                      high density filters, each containing 18,432 clones
                      (doubly spotted), represent the whole library for colony
                      screening and can be requested from the Clemson University
                      BAC/EST Resource Center (www.genome.clemson.edu)."
BASE COUNT       147 a     93 c    104 g    106 t      2 others
ORIGIN

  Query Match              78.2%;  Score 17.2;  DB 234;  Length 452;
  Best Local Similarity    86.4%;  Pred. No. 3.7e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 cagccgacggtagctgcgtcct 22
           ||  |||||| ||||||||||
Db     423 CAAGCGACGGGAGCTGCGTCCT 444


RESULT   5
BG365834
LOCUS       BG365834      481 bp    mRNA            EST       08-MAR-2001
DEFINITION  HVSMEi0004G14f Hordeum vulgare 20 DAP spike EST library HVcDNA0010
            (20 DAP) Hordeum vulgare cDNA clone HVSMEi0004G14f, mRNA sequence.
ACCESSION   BG365834
VERSION     BG365834.1  GI:13254933
KEYWORDS    EST.
SOURCE      barley.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 481)
  AUTHORS   Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu
            ,Y., Anderson,H., Dale,J., Henry,D., Kernodle,S., Palmer,M., Rambo
            ,T., Saski,C., Schwartzbeck,J., Simmons,J., Choi,D.W., Main,D. and
            Wood,T.
  TITLE     Development of a genetically and physically anchored EST resource
            for barley genomics
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: AATTAACCCTCACTAAAGGG
            High quality sequence stop: 464.
FEATURES              Location/Qualifiers
     source           1. .481
                      /organism="Hordeum vulgare"
                      /cultivar="Morex"
                      /db_xref="taxon:4513"
                      /clone="HVSMEi0004G14f"
                      /clone_lib="Hordeum vulgare 20 DAP spike EST library
                      HVcDNA0010 (20 DAP)"
                      /tissue_type="20 DAP spike"
                      /lab_host="SOLR"
                      /note="Vector: lambdaZAP; Site_1: EcoR1; Site_2: Xho1; For
                      more details on library preparation and sequence analysis
                      see http://www.genome.clemson.edu/projects/barley/"
BASE COUNT       74 a    145 c    198 g     64 t
ORIGIN

  Query Match              78.2%;  Score 17.2;  DB 152;  Length 481;
  Best Local Similarity    86.4%;  Pred. No. 3.7e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 cagccgacggtagctgcgtcct 22
           | |||||||||||| ||||| |
Db     440 CCGCCGACGGTAGCGGCGTCAT 461


RESULT   6
AI392116
LOCUS       AI392116      540 bp    mRNA            EST       03-FEB-1999
DEFINITION  NCC2B4T3 Conidial Neurospora crassa cDNA clone NC2B4 5' similar to
            ORF YNL212w, hypothetical 88.8 kD protein in RAP1-MER1 intergenic
            region, mRNA sequence.
ACCESSION   AI392116
VERSION     AI392116.1  GI:4219923
KEYWORDS    EST.
SOURCE      Neurospora crassa.
  ORGANISM  Neurospora crassa
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariales; Sordariaceae; Neurospora.
REFERENCE   1  (bases 1 to 540)
  AUTHORS   Nelson,M.A., Kang,S., Braun,E.L., Crawford,M.E., Dolan,P.L.,
            Leonard,P.M., Mitchell,J., Armijo,A.M., Bean,L., Blueyes,E.,
            Cushing,T., Errett,A., Fleharty,M., Gorman,M., Judson,K., Miller,R.
            , Ortega,J., Pavlova,I., Perea,J., Todisco,S., Trujillo,R.,
            Valentine,J., Wells,A., Werner-Washburne,M., Yazzie,S. and Natvig
            ,D.O.
  TITLE     Expressed sequences from conidial, mycelial, and sexual stages of
            Neurospora crassa
  JOURNAL   Fungal Genet. Biol. 21, 348-363 (1997)
  MEDLINE   97435549
COMMENT     Contact: Natvig,D.O./Nelson,M.A.
            Department of Biology
            University of New Mexico
            Castetter Hall, Albuquerque, NM 87131, USA
            Tel: 505 277 3411
            Fax: 505 277 0304
            Email: ngp@biology.unm.edu.
FEATURES              Location/Qualifiers
     source           1. .540
                      /organism="Neurospora crassa"
                      /strain="74-OR23-IV A"
                      /db_xref="taxon:5141"
```

```
                     /clone="NC2B4"
                     /clone_lib="Conidial"
                     /sex="Mating type A"
                     /tissue_type="Conidia"
                     /dev_stage="germinating conidia"
                     /note="mRNA isolated from germinating conidia, grown in 1x
                     Vogel's, 2% sucrose for 4.5 hours.  cDNA directionally
                     cloned into pBluescript SK(-) using the Uni-ZAP XR vector
                     system (Stratagene, La Jolla, CA)."
BASE COUNT      117 a    172 c    130 g    121 t
ORIGIN


  Query Match             78.2%;  Score 17.2;  DB 19;  Length 540;
  Best Local Similarity   86.4%;  Pred. No. 3.7e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcgtcct 22
          | ||||||||| |||| |||||
Db    309 CTGCCGACGGTCGCTGGGTCCT 330


RESULT   7
AW861895
LOCUS       AW861895      594 bp    mRNA            EST       19-MAY-2000
DEFINITION  CM3-CT0341-170200-093-f02 CT0341 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW861895
VERSION     AW861895.1  GI:7957588
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 594)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=CM3-CT0341-170
            200-093-f02&t3=2000-02-17&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 592.
FEATURES             Location/Qualifiers
     source          1. .594
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="CT0341"
                     /dev_stage="Adult"
                     /note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
[redacted]SE COUNT      134 a    161 c    162 g    137 t
[redacted]GIN


  Query Match             78.2%;  Score 17.2;  DB 121;  Length 594;
  Best Local Similarity   86.4%;  Pred. No. 3.7e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcgtcct 22
          ||||| |||||||||| | ||
Db    157 CAGCCCACGGTAGCTGCTTTCT 178


RESULT   8
AA950463
LOCUS       AA950463      744 bp    mRNA            EST       25-NOV-1998
DEFINITION  LD30267.5prime LD Drosophila melanogaster embryo pOT2 Drosophila
            melanogaster cDNA clone LD30267 5prime, mRNA sequence.
ACCESSION   AA950463
VERSION     AA950463.1  GI:3112700
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 744)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C., Brokstein
            ,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 302  row: F  column: 7
            High quality sequence stop: 623.
FEATURES             Location/Qualifiers
     source          1. .744
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="LD30267"
                     /clone_lib="LD Drosophila melanogaster embryo pOT2"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="XL1 Blue"
                     /note="Organ: embryo; Vector: pOT2; Site_1: EcoRI; Site_2:
                     XhoI; Sized fractionated cDNAs were directly ligated into
                     pOT2. "
BASE COUNT      186 a    238 c    197 g    123 t
ORIGIN


  Query Match             78.2%;  Score 17.2;  DB 14;  Length 744;
  Best Local Similarity   86.4%;  Pred. No. 3.7e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcgtcct 22
          |||||| ||| ||||| |||||
Db    575 CAGCCGCCGGGAGCTGAGTCCT 596


RESULT   9
BG414942
LOCUS       BG414942      910 bp    mRNA            EST       13-MAR-2001
DEFINITION  HVSMEk0004G02f Hordeum vulgare testa/pericarp EST library
            HVcDNA0013 (normal) Hordeum vulgare cDNA clone HVSMEk0004G02f, mRNA
            sequence.
ACCESSION   BG414942
VERSION     BG414942.1  GI:13320493
KEYWORDS    EST.
SOURCE      barley.
  ORGANISM  Hordeum vulgare
```

```
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 910)
  AUTHORS   Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu
            ,Y., Anderson,H., Dale,J., Henry,D., Kernodle,S., Palmer,M., Rambo
            ,T., Saski,C., Schwartzbeck,J., Simmons,J., Choi,D.W., Main,D. and
            Wood,T.
  TITLE     Development of a genetically and physically anchored EST resource
            for barley genomics
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: AATTAACCCTCACTAAAGGG
            High quality sequence stop: 799.
FEATURES             Location/Qualifiers
     source          1. .910
                     /organism="Hordeum vulgare"
                     /cultivar="Morex"
                     /db_xref="taxon:4513"
                     /clone="HVSMEk0004G02f"
                     /clone_lib="Hordeum vulgare testa/pericarp EST library
                     HVcDNA0013 (normal)"
                     /tissue_type="testa/pericarp"
                     /lab_host="TJC121"
                     /note="Vector: lambdaZAP; Site_1: EcoR1; Site_2: Xho1; For
                     more details on library preparation and sequence analysis
                     see http://www.genome.clemson.edu/projects/barley/"
BASE COUNT      191 a    243 c    289 g    184 t      3 others
ORIGIN


  Query Match             78.2%;  Score 17.2;  DB 153;  Length 910;
  Best Local Similarity   86.4%;  Pred. No. 3.6e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcgtcct 22
          | ||||||||||||| ||||| |
Db    456 CCGCCGACGGTAGCGGCGTCAT 477


RESULT  10
BE962535/c
LOCUS       BE962535      958 bp    mRNA            EST       14-DEC-2000
DEFINITION  601655848R1 NIH_MGC_66 Homo sapiens cDNA clone IMAGE:3855970 3',
            mRNA sequence.
ACCESSION   BE962535
VERSION     BE962535.2  GI:11765530
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 958)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     On Oct 3, 2000 this sequence version replaced gi:10573240.
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM570  row: o  column: 11
            High quality sequence stop: 101.
FEATURES             Location/Qualifiers
     source          1. .958
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3855970"
                     /clone_lib="NIH_MGC_66"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: ovary; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.8 kb.   Library constructed by Life
                     Technologies.  "
BASE COUNT      215 a    224 c    226 g    293 t
ORIGIN


  Query Match             78.2%;  Score 17.2;  DB 142;  Length 958;
  Best Local Similarity   86.4%;  Pred. No. 3.6e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcgtcct 22
          | ||||||||||||| ||||| |
Db    235 CAGCCGAGCGTGGCTGCGTCCT 214


RESULT  11
CNS05ORJ/c
LOCUS       CNS05ORJ      993 bp    DNA             GSS       26-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence SP6 end of clone
            006015 of library B from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL347032
VERSION     AL347032.1  GI:8240802
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 993)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 993)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 993)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .993
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="006015"
                     /clone_lib="B"
                     /note="Genoscope sequence ID : C0AB006AH08B1-end : SP6"
BASE COUNT      153 a    296 c    323 g    195 t     26 others
ORIGIN
```

```
  Query Match             78.2%;  Score 17.2;  DB 222;  Length 993;
  Best Local Similarity   86.4%;  Pred. No. 3.6e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 cagccgacggtagctgcgtcct 22
           |||| || || ||||||||||
Db     300 CAGCAGATGGCAGCTGCGTCCT 279


RESULT  12
AW265763
LOCUS       AW265763      259 bp    mRNA            EST       28-DEC-1999
DEFINITION  xp81e09.x1 NCI_CGAP_Ov40 Homo sapiens cDNA clone IMAGE:2746792 3'
            similar to TR:Q92584 Q92584 MYELOBLAST KIAA0198 ;, mRNA sequence.
ACCESSION   AW265763
VERSION     AW265763.1  GI:6642579
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 259)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Andrew Berchuck M.D., John Gillespie M.D.,
            Michael R. Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Trace considered overall poor quality
            Possible reversed clone: similarity on wrong strand
            Seq primer: -40UP from Gibco
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .259
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2746792"
                     /clone_lib="NCI_CGAP_Ov40"
                     /sex="female"
                     /tissue_type="endometrioid ovarian metastasis"
                     /lab_host="DH10B"
                     /note="Organ: ovary; Vector: pAMP10; cDNA made by oligo-dT
                     priming.  Non-directionally cloned into the UDG sites of
                     pAMP10.  Size-selected on agarose gel, average insert
                     size 500 bp.  Primary library; non-amplified.  cDNA
                     Library Preparation: David B. Krizman, Ph.D (NCI).
                     Reference:  Krizman et al. (1996) Cancer Research
                     56:5380-5383."
BASE COUNT       50 a     84 c     59 g     66 t
ORIGIN


  Query Match             76.4%;  Score 16.8;  DB 113;  Length 259;
  Best Local Similarity   90.0%;  Pred. No. 5.8e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 agccgacggtagctgcgtcc 21
           ||||| |||||||||||  ||
Db     126 AGCCGGCGGTAGCTGCGGCC 145


RESULT  13
BF825637
LOCUS       BF825637      281 bp    mRNA            EST       13-JAN-2001
DEFINITION  RC0-HN0024-161100-021-f06 HN0024 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF825637
VERSION     BF825637.1  GI:12168173
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 281)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=RC0&t2=RC0-HN0024-
            161100-021-f06&t3=2000-11-16&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 54
            High quality sequence stop: 281.
FEATURES             Location/Qualifiers
     source          1. .281
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="HN0024"
                     /dev_stage="Adult"
                     /note="Organ: head_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       56 a     88 c     80 g     57 t
ORIGIN


  Query Match             76.4%;  Score 16.8;  DB 170;  Length 281;
  Best Local Similarity   90.0%;  Pred. No. 5.8e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 agccgacggtagctgcgtcc 21
           |||| |||||||||| ||||
Db      35 AGCCCACGGTAGCTGTGTCC 54


RESULT  14
AU032424/c
LOCUS       AU032424      332 bp    mRNA            EST       20-OCT-1998
DEFINITION  AU032424 Rice root Oryza sativa cDNA clone R4067_1A, mRNA sequence.
ACCESSION   AU032424
VERSION     AU032424.1  GI:3768397
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
```

```
REFERENCE   1  (bases 1 to 332)
  AUTHORS   Minobe,Y. and Sasaki,T.
  TITLE     Rice cDNA from root
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305
            Tel: 0298-38-7441
            Fax: 0298-38-7468
            Email: tsasaki@abr.affrc.go.jp
            PROJECT ='RGP'.
FEATURES             Location/Qualifiers
     source          1. .332
                     /organism="Oryza sativa"
                     /strain="Nipponbare, sub_species Japonica"
                     /db_xref="taxon:4530"
                     /clone="R4067_1A"
                     /clone_lib="Rice root"
                     /note="Prepared from seedling root. "
BASE COUNT       47 a    121 c    105 g     51 t      8 others
ORIGIN
```

```
  Query Match             76.4%;  Score 16.8;  DB 107;  Length 332;
  Best Local Similarity   85.7%;  Pred. No. 5.8e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcgtcc 21
          ||| ||||||| || ||||||
Db    242 CAGNCGACGGTGGCGGCGTCC 222
```

```
RESULT  15
BF978349/c
LOCUS       BF978349      928 bp    mRNA            EST       22-JAN-2001
DEFINITION  602148933F2 NIH_MGC_62 Homo sapiens cDNA clone IMAGE:4307625 5',
            mRNA sequence.
ACCESSION   BF978349
VERSION     BF978349.1  GI:12345564
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 928)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1178  row: b  column: 10
            High quality sequence stop: 104.
FEATURES             Location/Qualifiers
     source          1. .928
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4307625"
                     /clone_lib="NIH_MGC_62"
                     /tissue_type="melanotic melanoma, high MDR"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: skin; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc);
                     Double-stranded cDNA was prepared from cell line RNA.  5'
                     and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).   Average
                     insert size 1.75 kb (range 0.9-4.0 kb).  15/15 colonies
                     contained  inserts by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT      235 a    319 c    247 g    127 t
ORIGIN
```

```
  Query Match             74.5%;  Score 16.4;  DB 172;  Length 928;
  Best Local Similarity   94.4%;  Pred. No. 8.7e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcg 18
          || ||||||||||||||||
Db    602 CAACCGACGGTAGCTGCG 585
```

```
Search completed: September 20, 2001, 23:28:18
Job time: 25558 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: September 21, 2001, 02:16:34 ; Search time 164.23 Seconds
(without alignments)
25.360 Million cell updates/sec

Title: US-09-138-735-9
Perfect score: 22
Sequence: 1 cagccgacggtagctgcgtcct 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn1_7/ptodata/1/ina/5A_COMB.seq:*
2: /cgn1_7/ptodata/1/ina/5B_COMB.seq:*
3: /cgn1_7/ptodata/1/ina/6A_COMB.seq:*
4: /cgn1_7/ptodata/1/ina/6B_COMB.seq:*
5: /cgn1_7/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn1_7/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	22	100.0	22	1	US-08-480-917-9	Sequence 9, Appli
	2	22	100.0	3402	1	US-08-480-917-1	Sequence 1, Appli
	3	15.6	70.9	269	4	US-08-482-918-58	Sequence 58, Appl
	4	15.6	70.9	269	4	US-09-224-681-58	Sequence 58, Appl
	5	15.6	70.9	269	4	US-08-336-728A-58	Sequence 58, Appl
	6	15.2	69.1	139	3	US-08-471-914-8	Sequence 8, Appli
c	7	15.2	69.1	790	2	US-08-715-204-2	Sequence 2, Appli
c	8	15.2	69.1	790	3	US-09-162-597-2	Sequence 2, Appli
c	9	15.2	69.1	2343	2	US-09-018-760-3	Sequence 3, Appli
	10	15.2	69.1	4060	1	US-08-308-949A-1	Sequence 1, Appli
	11	15.2	69.1	16885	1	US-08-390-878-16	Sequence 16, Appl
	12	15.2	69.1	80161	3	US-09-036-987A-1	Sequence 1, Appli
	13	14.8	67.3	788	4	US-08-998-416-440	Sequence 440, App
c	14	14.8	67.3	1894	4	US-07-912-122-3	Sequence 3, Appli
c	15	14.8	67.3	1894	5	PCT-US93-06404-3	Sequence 3, Appli
c	16	14.8	67.3	3453	1	US-07-813-593-1	Sequence 1, Appli
c	17	14.8	67.3	3453	1	US-07-977-451-1	Sequence 1, Appli
c	18	14.8	67.3	3453	1	US-07-946-507-1	Sequence 1, Appli
c	19	14.8	67.3	3453	1	US-08-252-517-1	Sequence 1, Appli
c	20	14.8	67.3	3453	1	US-07-906-397A-1	Sequence 1, Appli
c	21	14.8	67.3	3453	1	US-08-601-891-1	Sequence 1, Appli
c	22	14.8	67.3	3453	2	US-09-021-324-1	Sequence 1, Appli
c	23	14.8	67.3	3453	5	PCT-US92-02750-1	Sequence 1, Appli
c	24	14.8	67.3	3453	5	PCT-US92-05401-1	Sequence 1, Appli
c	25	14.8	67.3	3453	5	PCT-US92-09893-1	Sequence 1, Appli
c	26	14.8	67.3	3521	1	US-08-222-299-1	Sequence 11, Appl
c	27	14.8	67.3	3521	2	US-08-434-878-1	Sequence 1, Appli
c	28	14.8	67.3	3521	5	PCT-US95-03718-1	Sequence 1, Appli
c	29	14.6	66.4	348	2	US-08-610-728B-3	Sequence 3, Appli
c	30	14.6	66.4	348	2	US-08-610-728B-11	Sequence 11, Appl
c	31	14.6	66.4	1207	2	US-08-525-864A-5	Sequence 5, Appli
c	32	14.6	66.4	1320	1	US-08-176-412-1	Sequence 1, Appli
c	33	14.6	66.4	1320	2	US-08-555-268A-1	Sequence 1, Appli
c	34	14.6	66.4	1320	2	US-08-495-695B-1	Sequence 1, Appli
c	35	14.6	66.4	1320	5	PCT-US94-14436-1	Sequence 1, Appli
c	36	14.6	66.4	1596	1	US-08-531-601-2	Sequence 2, Appli
c	37	14.6	66.4	1596	2	US-08-859-032-2	Sequence 2, Appli
c	38	14.6	66.4	1607	3	US-08-753-007A-3	Sequence 3, Appli
c	39	14.6	66.4	1607	3	US-09-398-496-3	Sequence 3, Appli
c	40	14.6	66.4	2467	3	US-08-753-007A-1	Sequence 1, Appli
c	41	14.6	66.4	2467	3	US-09-398-496-1	Sequence 1, Appli
c	42	14.6	66.4	3441	2	US-08-525-864A-1	Sequence 1, Appli
c	43	14.6	66.4	3825	1	US-08-737-597-1	Sequence 1, Appli
c	44	14.6	66.4	3825	1	US-08-737-597-2	Sequence 2, Appli
c	45	14.6	66.4	6312	1	US-08-531-601-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-480-917-9
; Sequence 9, Application US/08480917
; Patent No. 5820864
; GENERAL INFORMATION:
; APPLICANT: PARANHOS-BACCALA, Glaucia
; APPLICANT: LESENECHAL, Mylene
; APPLICANT: JOLIVET, Michel
; TITLE OF INVENTION: NEW TRYPANOSOMA CRUZI ANTIGEN, AND GENE
; TITLE OF INVENTION: ENCODING THE LATTER; THEIR APPLICATION TO THE DETECTION OF
; TITLE OF INVENTION: CHAGAS DISEASE
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oliff & Berridge
; STREET: 700 South Washington Street, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,917
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 36400
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 22 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-480-917-9

Query Match 100.0%; Score 22; DB 1; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cagccgacggtagctgcgtcct 22

```
          ||||||||||||||||||||
Db      1 CAGCCGACGGTAGCTGCGTCCT 22


RESULT   2
US-08-480-917-1
; Sequence 1, Application US/08480917
; Patent No. 5820864
;  GENERAL INFORMATION:
;    APPLICANT:  PARANHOS-BACCALA, Glaucia
;    APPLICANT:  LESENECHAL, Mylene
;    APPLICANT:  JOLIVET, Michel
;    TITLE OF INVENTION:  NEW TRYPANOSOMA CRUZI ANTIGEN, AND GENE
;    TITLE OF INVENTION:  ENCODING THE LATTER; THEIR APPLICATION TO THE DETECTION OF
;    TITLE OF INVENTION:  CHAGAS DISEASE
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Oliff & Berridge
;      STREET:  700 South Washington Street, Suite 300
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22314
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,917
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berridge, William P.
;      REGISTRATION NUMBER:  30,024
;      REFERENCE/DOCKET NUMBER:  WPB 36400
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-836-6400
;      TELEFAX:  703-836-2787
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3402 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-480-917-1


  Query Match             100.0%;  Score 22;  DB 1;  Length 3402;
  Best Local Similarity   100.0%;  Pred. No. 0.13;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcgtcct 22
          ||||||||||||||||||||||
Db   1266 CAGCCGACGGTAGCTGCGTCCT 1287


RESULT   3
US-08-482-918-58
; Sequence 58, Application US/08482918
; Patent No. 6207417
;  GENERAL INFORMATION:
;    APPLICANT:  Zsebo, Krisztina M.
;    APPLICANT:  Bosselman, Robert A.
;    APPLICANT:  Suggs, Sidney V.
;    APPLICANT:  Martin, Francis H.
;    TITLE OF INVENTION:  Stem Cell Factor
;    NUMBER OF SEQUENCES:  104
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/482,918
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  01017/33005
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  58:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  269 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(1..210, 223..258)
US-08-482-918-58


  Query Match             70.9%;  Score 15.6;  DB 4;  Length 269;
  Best Local Similarity   81.8%;  Pred. No. 89;
  Matches   18;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcgtcct 22
          | |||| |||| ||||| ||||
Db    139 CTGCCGCCGGTTGCTGCTTCCT 160


RESULT   4
US-09-224-681-58
; Sequence 58, Application US/09224681
; Patent No. 6207454
;  GENERAL INFORMATION:
;    APPLICANT:  Zsebo, Krisztina M.
;    APPLICANT:  Bosselman, Robert A.
;    APPLICANT:  Suggs, Sidney V.
;    APPLICANT:  Martin, Francis H.
;    TITLE OF INVENTION:  Method for Enhancing the Efficiency of Gene
;    TITLE OF INVENTION:  Transfer with Stem Cell Factor (SCF) Polypeptide
;    NUMBER OF SEQUENCES:  104
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/224,681
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  09/005,893
;      FILING DATE:  12-JAN-1998
```

```
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/449,653
;       FILING DATE:  24-MAY-1995
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/982,255
;       FILING DATE:  25-NOV-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/589,701
;       FILING DATE:  01-OCT-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/573,616
;       FILING DATE:  24-AUG-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/537,198
;       FILING DATE:  11-JUN-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/422,383
;       FILING DATE:  16-OCT-1989
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Clough, David W.
;       REGISTRATION NUMBER:  36,107
;       REFERENCE/DOCKET NUMBER:  01017/35199
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  312/474-6300
;       TELEFAX:  312/474-0448
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  58:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  269 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  join(1..210, 223..258)
US-09-224-681-58


  Query Match             70.9%;  Score 15.6;  DB 4;  Length 269;
  Best Local Similarity   81.8%;  Pred. No. 89;
  Matches   18;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcgtcct 22
          | |||| |||| ||||| ||||
Db    139 CTGCCGCCGGTTGCTGCTTCCT 160


RESULT   5
US-08-336-728A-58
; Sequence 58, Application US/08336728A
; Patent No. 6207802
;  GENERAL INFORMATION:
;    APPLICANT:  Zsebo, Krisztina M.
;    APPLICANT:  Bosselman, Robert A.
;    APPLICANT:  Suggs, Sidney V.
;    APPLICANT:  Martin, Francis H.
;    TITLE OF INVENTION:  Stem Cell Factor
;    NUMBER OF SEQUENCES:  104
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/336,728A
;      FILING DATE:  09-NOV-1994
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/982,255
;      FILING DATE:  25-NOV-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/589,701
;      FILING DATE:  01-OCT-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/573,616
;      FILING DATE:  24-AUG-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/537,198
;      FILING DATE:  11-JUN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/422,383
;      FILING DATE:  16-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  01017/32956
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  58:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  269 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(1..210, 223..258)
US-08-336-728A-58


  Query Match             70.9%;  Score 15.6;  DB 4;  Length 269;
  Best Local Similarity   81.8%;  Pred. No. 89;
  Matches   18;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcgtcct 22
          | |||| |||| ||||| ||||
Db    139 CTGCCGCCGGTTGCTGCTTCCT 160


RESULT   6
US-08-471-914-8
; Sequence 8, Application US/08471914A
; Patent No. 6057152
; GENERAL INFORMATION:
;  APPLICANT: Samulski, R.
;  APPLICANT: Xiao, X.
;  TITLE OF INVENTION: RECOMBINANT VIRAL VECTOR SYSTEM
;  FILE REFERENCE: 6636-027
;  CURRENT APPLICATION NUMBER: US/08/471,914A
;  CURRENT FILING DATE: 1995-06-06
;  EARLIER APPLICATION NUMBER: 08/440,738
;  EARLIER FILING DATE: 1995-05-15
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
;   LENGTH: 139
;   TYPE: DNA
;   ORGANISM: adeno-associated virus
US-08-471-914-8


  Query Match             69.1%;  Score 15.2;  DB 3;  Length 139;
  Best Local Similarity   85.0%;  Pred. No. 1.3e+02;
```

```
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 gccgacggtagctgcgtcct 22
          |||||||||| | ||| |||
Db     95 gccgacggtatcagcgccct 114


RESULT   7
US-08-715-204-2/c
; Sequence 2, Application US/08715204
; Patent No. 5874286
;  GENERAL INFORMATION:
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Au-Young, Janice
;    APPLICANT:  Goli, Surya K.
;    APPLICANT:  Hillman, Jennifer.
;    APPLICANT:  Zweiger, Gary B.
;    TITLE OF INVENTION:  A NOVEL TUMOR PROTEIN
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  U.S.
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/715,204
;      FILING DATE:  Filed Herewith
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0126 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  790 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    IMMEDIATE SOURCE:
;      LIBRARY:
;      CLONE:  Consensus
US-08-715-204-2


  Query Match             69.1%;  Score 15.2;  DB 2;  Length 790;
  Best Local Similarity   85.0%;  Pred. No. 1.3e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 agccgacggtagctgcgtcc 21
          ||| || ||||||||||  ||
Db     42 AGCAGATGGTAGCTGCGCCC 23


RESULT   8
US-09-162-597-2/c
; Sequence 2, Application US/09162597
; Patent No. 6043343
;  GENERAL INFORMATION:
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Au-Young, Janice
;    APPLICANT:  Goli, Surya K.
;    APPLICANT:  Hillman, Jennifer.
;    APPLICANT:  Zweiger, Gary B.
;    TITLE OF INVENTION:  A NOVEL TUMOR PROTEIN
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  U.S.
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/162,597
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/715,204
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0126 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  790 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    IMMEDIATE SOURCE:
;      LIBRARY:
;      CLONE:  Consensus
US-09-162-597-2


  Query Match             69.1%;  Score 15.2;  DB 3;  Length 790;
  Best Local Similarity   85.0%;  Pred. No. 1.3e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 agccgacggtagctgcgtcc 21
          ||| || ||||||||||  ||
Db     42 AGCAGATGGTAGCTGCGCCC 23


RESULT   9
US-09-018-760-3/c
; Sequence 3, Application US/09018760
; Patent No. 5981249
;  GENERAL INFORMATION:
;    APPLICANT:  Shi, Qinwei
;    APPLICANT:  Tobias, Rowel
;    TITLE OF INVENTION:  A SINGLE CHAIN CREATINE KINASE MB
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Klauber & Jackson
;      STREET:  411 Hackensack Avenue, 4th Floor
;      CITY:  Hackensack
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/018,760
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jackson Esq., David A.
;      REGISTRATION NUMBER:  26,742
;      REFERENCE/DOCKET NUMBER:  1112-1-050
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-487-5800
;      TELEFAX:  201-343-1684
;      TELEX:  133521
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2343 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
US-09-018-760-3


  Query Match             69.1%;  Score 15.2;  DB 2;  Length 2343;
  Best Local Similarity   85.0%;  Pred. No. 1.3e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 agccgacggtagctgcgtcc 21
           |||| |||| ||||| ||||
Db     994 AGCCCACGGCAGCTGTGTCC 975


RESULT  10
US-08-308-949A-1
; Sequence 1, Application US/08308949A
; Patent No. 5580703
;  GENERAL INFORMATION:
;    APPLICANT:  Kotin, Robert M.
;    APPLICANT:  Berns, Kenneth I.
;    APPLICANT:  Linden, Ralph M.
;    TITLE OF INVENTION:  Human Adeno-Associated Virus Integration
;    TITLE OF INVENTION:  Site DNA and Uses Thereof
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/308,949A
;      FILING DATE:  September 20, 1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/947,127
;      FILING DATE:  September 27, 1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Alice O.
;      REGISTRATION NUMBER:  33,542
;      REFERENCE/DOCKET NUMBER:  ACC92-10F
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-861-6240
;      TELEFAX:  617-861-9540
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4060 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-308-949A-1


  Query Match             69.1%;  Score 15.2;  DB 1;  Length 4060;
  Best Local Similarity   85.0%;  Pred. No. 1.3e+02;,
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 gccgacggtagctgcgtcct 22
           |||||||||| | ||| |||
Db     721 GCCGACGGTATCAGCGCCCT 740


RESULT  11
US-08-390-878-16
; Sequence 16, Application US/08390878
; Patent No. 5700683
;  GENERAL INFORMATION:
;    APPLICANT:  Stover, Charles K.
;    APPLICANT:  Mahairas, Gregory G.
;    TITLE OF INVENTION:  VIRULENCE-ATTENUATING GENETIC DELETIONS
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend Khourie and Crew
;      STREET:  One Market Plaza, Steuart Street Tower, 20th
;      STREET:  Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/390,878
;      FILING DATE:  17-FEB-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hunter, Tom
;      REGISTRATION NUMBER:  38,498
;      REFERENCE/DOCKET NUMBER:  15371A-17
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/543/9600
;      TELEFAX:  415/543/5043
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  16885 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-390-878-16


  Query Match             69.1%;  Score 15.2;  DB 1;  Length 16885;
  Best Local Similarity   85.0%;  Pred. No. 1.3e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 cagccgacggtagctgcgtc 20
           |||||||||  ||||||| |
Db   12302 CAGCCGACGTCAGCTGCGCC 12321


RESULT  12
US-09-036-987A-1
; Sequence 1, Application US/09036987A
; Patent No. 6143526
;  GENERAL INFORMATION:
;    APPLICANT:  Baltz, Richard H.
```

```
;    APPLICANT:  Broughton, Mary C.
;    APPLICANT:  Crawford, Kathryn P.
;    APPLICANT:  Madduri, Krishnamurthy
;    APPLICANT:  Merlo, Donald J.
;    APPLICANT:  Treadway, Patti J.
;    APPLICANT:  Turner, Jan R.
;    APPLICANT:  Waldron, Clive
;    TITLE OF INVENTION:  Biosynthetic Genes For Spinosyn Insecticide
;    TITLE OF INVENTION:  Production
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dow AgroSciences LLC Patent Department
;      STREET:  9330 Zionsville Road
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  USA
;      ZIP:  46268
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/036,987A
;      FILING DATE:  09-MAR-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stuart, Donald R
;      REGISTRATION NUMBER:  28,479
;      REFERENCE/DOCKET NUMBER:  50,608
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317)337-4816
;      TELEFAX:  (317)337-4847
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  80161 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-036-987A-1


  Query Match             69.1%;  Score 15.2;  DB 3;  Length 80161;
  Best Local Similarity   85.0%;  Pred. No. 1.2e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 cagccgacggtagctgcgtc 20
           || ||| ||||||| |||||
Db   12603 CACCCGGCGGTAGCGGCGTC 12622


RESULT  13
US-08-998-416-440
; Sequence 440, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
;    APPLICANT:  Philippsen, Peter
;    APPLICANT:  Pohlmann, Rainer
;    APPLICANT:  Steiner, Sabine
;    APPLICANT:  Mohr, Christine
;    APPLICANT:  Wendland, Jurgen
;    APPLICANT:  Knechtle, Philipp
;    APPLICANT:  Rebischung, Corinne
;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  1152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6239264artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  No. 6239264th Carolina
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/998,416
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 0016/97
;      FILING DATE:  31-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  PF/5-30306/A/CGC1976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  440:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  788 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  PAG1321RP
US-08-998-416-440


  Query Match             67.3%;  Score 14.8;  DB 4;  Length 788;
  Best Local Similarity   88.9%;  Pred. No. 2e+02;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       5 cgacggtagctgcgtcct 22
           || ||||||| |||||||
Db   363 CGGCGGTAGCAGCGTCCT 380


RESULT  14
US-07-912-122-3/c
; Sequence 3, Application US/07912122
; Patent No. 6228609
;  GENERAL INFORMATION:
;    APPLICANT:  YANG, Zhi
;    TITLE OF INVENTION:  NOVEL FLK-2 AND ANALOGS
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bertram I. Rowland
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/912,122
;      FILING DATE:  19920709
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rowland, Bertram I.
;      REGISTRATION NUMBER:  20,015
;      REFERENCE/DOCKET NUMBER:  A-55931/BIR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415)789-1989
;      TELEFAX:  (415) 398-3249
;  INFORMATION FOR SEQ ID NO:  3:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1894 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  31..1473
US-07-912-122-3


  Query Match            67.3%;  Score 14.8;  DB 4;  Length 1894;
  Best Local Similarity  88.9%;  Pred. No. 2e+02;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcg 18
          |||||| |||| ||||||
Db     66 CAGCCGCCGGTCGCTGCG 49


RESULT  15
PCT-US93-06404-3/c
; Sequence 3, Application PC/TUS9306404
;  GENERAL INFORMATION:
;    APPLICANT:  YANG, Zhi
;    TITLE OF INVENTION:  NOVEL FLK-2 AND ANALOGS
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bertram I. Rowland
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/06404
;      FILING DATE:  19930707
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rowland, Bertram I.
;      REGISTRATION NUMBER:  20,015
;      REFERENCE/DOCKET NUMBER:  FP-55931/BIR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415)789-1989
;      TELEFAX:  (415) 398-3249
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1894 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  31..1473
PCT-US93-06404-3


  Query Match            67.3%;  Score 14.8;  DB 5;  Length 1894;
  Best Local Similarity  88.9%;  Pred. No. 2e+02;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcg 18
          |||||| |||| ||||||
Db     66 CAGCCGCCGGTCGCTGCG 49
```

```
Search completed: September 21, 2001, 02:16:39
Job time: 28232 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: September 21, 2001, 02:19:48 ; Search time 6407.28 Seconds
(without alignments)
50.738 Million cell updates/sec

Title: US-09-138-735-9
Perfect score: 22
Sequence: 1 cagccgacggtagctgcgtcct 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 14155048 seqs, 7388405095 residues

Total number of hits satisfying chosen parameters: 28310096

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*
1: /cgn1_7/ptodata/1/pna/PCTUS_COMB.seq:*
2: /cgn1_7/ptodata/1/pna/US06_COMB.seq:*
3: /cgn1_7/ptodata/1/pna/US07_COMB.seq:*
4: /cgn1_7/ptodata/1/pna/US080_COMB.seq:*
5: /cgn1_7/ptodata/1/pna/US081_COMB.seq:*
6: /cgn1_7/ptodata/1/pna/US082_COMB.seq:*
7: /cgn1_7/ptodata/1/pna/US083_COMB.seq:*
8: /cgn1_7/ptodata/1/pna/US084_COMB.seq:*
9: /cgn1_7/ptodata/1/pna/US085_COMB.seq:*
10: /cgn1_7/ptodata/1/pna/US086_COMB.seq:*
11: /cgn1_7/ptodata/1/pna/US087_COMB.seq:*
12: /cgn1_7/ptodata/1/pna/US088_COMB.seq:*
13: /cgn1_7/ptodata/1/pna/US089_COMB.seq:*
14: /cgn1_7/ptodata/1/pna/US090_COMB.seq:*
15: /cgn1_7/ptodata/1/pna/US091_COMB.seq:*
16: /cgn1_7/ptodata/1/pna/US092_COMB.seq:*
17: /cgn1_7/ptodata/1/pna/US093_COMB.seq:*
18: /cgn1_7/ptodata/1/pna/US094_COMB.seq:*
19: /cgn1_7/ptodata/1/pna/US095A_COMB.seq:*
20: /cgn1_7/ptodata/1/pna/US095B_COMB.seq:*
21: /cgn1_7/ptodata/1/pna/US095C_COMB.seq:*
22: /cgn1_7/ptodata/1/pna/US095D_COMB.seq:*
23: /cgn1_7/ptodata/1/pna/US096A_COMB.seq:*
24: /cgn1_7/ptodata/1/pna/US096B_COMB.seq:*
25: /cgn1_7/ptodata/1/pna/US096C_COMB.seq:*
26: /cgn1_7/ptodata/1/pna/US096D_COMB.seq:*
27: /cgn1_7/ptodata/1/pna/US096E_COMB.seq:*
28: /cgn1_7/ptodata/1/pna/US097A_COMB.seq:*
29: /cgn1_7/ptodata/1/pna/US097B_COMB.seq:*
30: /cgn1_7/ptodata/1/pna/US097C_COMB.seq:*
31: /cgn1_7/ptodata/1/pna/US098_COMB.seq:*
32: /cgn1_7/ptodata/1/pna/US6000_COMB.seq:*
33: /cgn1_7/ptodata/1/pna/US6001_COMB.seq:*
34: /cgn1_7/ptodata/1/pna/US6002_COMB.seq:*
35: /cgn1_7/ptodata/1/pna/US6003_COMB.seq:*
36: /cgn1_7/ptodata/1/pna/US6004_COMB.seq:*
37: /cgn1_7/ptodata/1/pna/US6005_COMB.seq:*
38: /cgn1_7/ptodata/1/pna/US6006_COMB.seq:*
39: /cgn1_7/ptodata/1/pna/US6007_COMB.seq:*
40: /cgn1_7/ptodata/1/pna/US6008_COMB.seq:*
41: /cgn1_7/ptodata/1/pna/US6009_COMB.seq:*
42: /cgn1_7/ptodata/1/pna/US6010_COMB.seq:*
43: /cgn1_7/ptodata/1/pna/US6011_COMB.seq:*
44: /cgn1_7/ptodata/1/pna/US6012_COMB.seq:*
45: /cgn1_7/ptodata/1/pna/US6013_COMB.seq:*
46: /cgn1_7/ptodata/1/pna/US6014_COMB.seq:*
47: /cgn1_7/ptodata/1/pna/US6015_COMB.seq:*
48: /cgn1_7/ptodata/1/pna/US6016_COMB.seq:*
49: /cgn1_7/ptodata/1/pna/US6017_COMB.seq:*
50: /cgn1_7/ptodata/1/pna/US6018_COMB.seq:*
51: /cgn1_7/ptodata/1/pna/US6019_COMB.seq:*
52: /cgn1_7/ptodata/1/pna/US6020_COMB.seq:*
53: /cgn1_7/ptodata/1/pna/US6021_COMB.seq:*
54: /cgn1_7/ptodata/1/pna/US6022_COMB.seq:*
55: /cgn1_7/ptodata/1/pna/US6023_COMB.seq:*
56: /cgn1_7/ptodata/1/pna/US6024_COMB.seq:*
57: /cgn1_7/ptodata/1/pna/US6025_COMB.seq:*
58: /cgn1_7/ptodata/1/pna/US6026_COMB.seq:*
59: /cgn1_7/ptodata/1/pna/US6027_COMB.seq:*
60: /cgn1_7/ptodata/1/pna/US6028_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	22	100.0	22	13	US-08-988-242-5	Sequence 5, Appli
2	22	100.0	22	15	US-09-138-736-9	Sequence 9, Appli
3	22	100.0	3402	13	US-08-988-242-1	Sequence 1, Appli
4	22	100.0	3402	15	US-09-138-736-1	Sequence 1, Appli
c 5	17.8	80.9	465	16	US-09-270-849B-93552	Sequence 93552, A
c 6	17.2	78.2	87	16	US-09-270-849B-46264	Sequence 46264, A
7	17.2	78.2	253	16	US-09-244-000A-76417	Sequence 76417, A
8	17.2	78.2	264	25	US-09-654-617-106410	Sequence 106410,
9	17.2	78.2	264	27	US-09-684-016-106410	Sequence 106410,
10	17.2	78.2	425	25	US-09-654-617-109675	Sequence 109675,
11	17.2	78.2	425	27	US-09-684-016-109675	Sequence 109675,
c 12	17.2	78.2	426	16	US-09-270-849B-122012	Sequence 122012,
c 13	17.2	78.2	2339	49	US-60-173-464-23049	Sequence 23049, A
c 14	17.2	78.2	2339	51	US-60-191-637-28973	Sequence 28973, A
c 15	17.2	78.2	2339	51	US-60-191-681-23456	Sequence 23456, A
c 16	17.2	78.2	4133	24	US-09-620-392-11665	Sequence 11665, A
c 17	17.2	78.2	7215	24	US-09-620-392-9977	Sequence 9977, Ap
18	17.2	78.2	9348	24	US-09-620-392-25910	Sequence 25910, A
c 19	17.2	78.2	9520	24	US-09-620-392-30257	Sequence 30257, A
c 20	17.2	78.2	9848	28	US-09-702-134-5934	Sequence 5934, Ap
c 21	17.2	78.2	10869	24	US-09-620-392-568	Sequence 568, App
c 22	17.2	78.2	12654	24	US-09-620-392-32904	Sequence 32904, A
c 23	17.2	78.2	12654	28	US-09-702-134-25170	Sequence 25170, A
24	17.2	78.2	13891	24	US-09-620-392-65700	Sequence 65700, A
c 25	17.2	78.2	14396	24	US-09-620-392-58244	Sequence 58244, A
26	17.2	78.2	19687	24	US-09-620-392-13752	Sequence 13752, A
27	17.2	78.2	27082	1	PCT-US01-01354-25259	Sequence 25259, A
c 28	17.2	78.2	29326	24	US-09-620-392-48151	Sequence 48151, A
29	17.2	78.2	29595	49	US-60-171-627-625	Sequence 625, App
30	17.2	78.2	29595	49	US-60-173-464-6116	Sequence 6116, Ap
c 31	17.2	78.2	30905	28	US-09-702-134-7927	Sequence 7927, Ap
32	17.2	78.2	33078	19	US-09-528-237A-1201	Sequence 1201, Ap
c 33	17.2	78.2	34222	24	US-09-620-392-26868	Sequence 26868, A
34	17.2	78.2	35363	24	US-09-620-392-44557	Sequence 44557, A
c 35	17.2	78.2	38678	28	US-09-702-134-5608	Sequence 5608, Ap
36	17.2	78.2	38699	24	US-09-620-392-13645	Sequence 13645, A
37	17.2	78.2	57538	28	US-09-702-134-6237	Sequence 6237, Ap
c 38	16.8	76.4	402	18	US-09-436-762A-27139	Sequence 27139, A
c 39	16.8	76.4	527	16	US-09-270-849B-151280	Sequence 151280,
c 40	16.8	76.4	2540	52	US-60-207-216-209	Sequence 209, App
c 41	16.8	76.4	66915	48	US-60-161-932-304	Sequence 304, App
c 42	16.4	74.5	417	16	US-09-270-849B-56976	Sequence 56976, A
c 43	16.4	74.5	478	16	US-09-270-849B-121200	Sequence 121200,
c 44	16.4	74.5	498	16	US-09-270-849B-39425	Sequence 39425, A
c 45	16.4	74.5	107600	20	US-09-534-859-571	Sequence 571, App

ALIGNMENTS

```
RESULT   1
US-08-988-242-5
; Sequence 5, Application US/08988242
;  GENERAL INFORMATION:
;    APPLICANT:  PARANHOS-BACCALA, GLAUCIA
;    APPLICANT:  LESENECHAL, MYLENE
;    APPLICANT:  JOLIVET, MICHEL
;    APPLICANT:  MANDRAND, BERNARD
;    TITLE OF INVENTION:  NEW TRYPANOSOMA CRUZI ANTIGEN, GENE
;    TITLE OF INVENTION:  ENCODING THEREFOR, AND METHODS OF DETECTING AND TREATING
;    TITLE OF INVENTION:  CHAGAS DISEASE
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OLIFF & BERRIDGE, PLC
;      STREET:  P.O. BOX 19928
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22320
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/988,242
;      FILING DATE:  10-DEC-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berridge, William P.
;      REGISTRATION NUMBER:  30,024
;      REFERENCE/DOCKET NUMBER:  WPB 36400A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-836-6400
;      TELEFAX:  703-836-2787
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  22 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-988-242-5


  Query Match           100.0%;  Score 22;  DB 13;  Length 22;
  Best Local Similarity  100.0%;  Pred. No. 3.1;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcgtcct 22
          ||||||||||||||||||||||
Db      1 CAGCCGACGGTAGCTGCGTCCT 22


RESULT   2
US-09-138-736-9
; Sequence 9, Application US/09138736
;  GENERAL INFORMATION:
;    APPLICANT:  PARANHOS-BACCALA, Glaucia
;    APPLICANT:  LESENECHAL, Mylene
;    APPLICANT:  JOLIVET, Michel
;    TITLE OF INVENTION:  NEW TRYPANOSOMA CRUZI ANTIGEN, AND GENE
;    TITLE OF INVENTION:  ENCODING THE LATTER; THEIR APPLICATION TO THE DETECTION OF
;    TITLE OF INVENTION:  CHAGAS DISEASE
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Oliff & Berridge
;      STREET:  700 South Washington Street, Suite 300
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22314
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/138,736
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/480,917
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berridge, William P.
;      REGISTRATION NUMBER:  30,024
;      REFERENCE/DOCKET NUMBER:  WPB 36400
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-836-6400
;      TELEFAX:  703-836-2787
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  22 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-138-736-9


  Query Match           100.0%;  Score 22;  DB 15;  Length 22;
  Best Local Similarity  100.0%;  Pred. No. 3.1;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcgtcct 22
          ||||||||||||||||||||||
Db      1 CAGCCGACGGTAGCTGCGTCCT 22


RESULT   3
US-08-988-242-1
; Sequence 1, Application US/08988242
;  GENERAL INFORMATION:
;    APPLICANT:  PARANHOS-BACCALA, GLAUCIA
;    APPLICANT:  LESENECHAL, MYLENE
;    APPLICANT:  JOLIVET, MICHEL
;    APPLICANT:  MANDRAND, BERNARD
;    TITLE OF INVENTION:  NEW TRYPANOSOMA CRUZI ANTIGEN, GENE
;    TITLE OF INVENTION:  ENCODING THEREFOR, AND METHODS OF DETECTING AND TREATING
;    TITLE OF INVENTION:  CHAGAS DISEASE
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OLIFF & BERRIDGE, PLC
;      STREET:  P.O. BOX 19928
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22320
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/988,242
;      FILING DATE:  10-DEC-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berridge, William P.
;      REGISTRATION NUMBER:  30,024
;      REFERENCE/DOCKET NUMBER:  WPB 36400A
```

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-836-6400
;      TELEFAX:  703-836-2787
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3402 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-988-242-1


  Query Match            100.0%;  Score 22;  DB 13;  Length 3402;
  Best Local Similarity   100.0%;  Pred. No. 3.2;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcgtcct 22
          ||||||||||||||||||||||
Db   1266 CAGCCGACGGTAGCTGCGTCCT 1287


RESULT   4
US-09-138-736-1
; Sequence 1, Application US/09138736
;  GENERAL INFORMATION:
;    APPLICANT:  PARANHOS-BACCALA, Glaucia
;    APPLICANT:  LESENECHAL, Mylene
;    APPLICANT:  JOLIVET, Michel
;    TITLE OF INVENTION:  NEW TRYPANOSOMA CRUZI ANTIGEN, AND GENE
;    TITLE OF INVENTION:  ENCODING THE LATTER; THEIR APPLICATION TO THE DETECTION OF
;    TITLE OF INVENTION:  CHAGAS DISEASE
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Oliff & Berridge
;      STREET:  700 South Washington Street, Suite 300
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22314
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/138,736
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/480,917
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berridge, William P.
;      REGISTRATION NUMBER:  30,024
;      REFERENCE/DOCKET NUMBER:  WPB 36400
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-836-6400
;      TELEFAX:  703-836-2787
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3402 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-138-736-1


  Query Match            100.0%;  Score 22;  DB 15;  Length 3402;
  Best Local Similarity   100.0%;  Pred. No. 3.2;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcgtcct 22
          ||||||||||||||||||||||
Db   1266 CAGCCGACGGTAGCTGCGTCCT 1287


RESULT   5
US-09-270-849B-93552/c
; Sequence 93552, Application US/09270849B
; GENERAL INFORMATION:
;  APPLICANT: Swimmer et al.
;  TITLE OF INVENTION: Insect genome survey devices
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/09/270,849B
;  CURRENT FILING DATE:  1999-03-17
;  NUMBER OF SEQ ID NOS: 195450
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 93552
;   LENGTH: 465
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-270-849B-93552


  Query Match            80.9%;  Score 17.8;  DB 16;  Length 465;
  Best Local Similarity   90.5%;  Pred. No. 3.1e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 agccgacggtagctgcgtcct 22
          |||||||| ||||||||| ||
Db    375 AGCCGACGCTAGCTGCGTTCT 355


RESULT   6
US-09-270-849B-46264/c
; Sequence 46264, Application US/09270849B
; GENERAL INFORMATION:
;  APPLICANT: Swimmer et al.
;  TITLE OF INVENTION: Insect genome survey devices
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/09/270,849B
;  CURRENT FILING DATE:  1999-03-17
;  NUMBER OF SEQ ID NOS: 195450
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46264
;   LENGTH: 87
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-270-849B-46264


  Query Match            78.2%;  Score 17.2;  DB 16;  Length 87;
  Best Local Similarity   86.4%;  Pred. No. 5.8e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcgtcct 22
          |||||||||||| |||| | ||
Db     33 CAGCCGACGGTACCTGCATTCT 12


RESULT   7
US-09-244-000A-76417
; Sequence 76417, Application US/09244000A
; GENERAL INFORMATION:
;  APPLICANT:       Byrum, Joseph R.
;  APPLICANT:  La Rosa, Thomas J.
;  APPLICANT:  Thompson, Michael D.
;  TITLE OF INVENTION:       Nucleic Acid Molecules and Other Molecules Associated wi
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE:       38-21(15305)B
```

```
;  CURRENT APPLICATION NUMBER: US/09/244,000A
;  CURRENT FILING DATE:  1999-02-08
;  PRIOR APPLICATION DATA removed - refer to PALM or File Wrapper
;  NUMBER OF SEQ ID NOS:       101193
; SEQ ID NO 76417
;   LENGTH: 253
;   TYPE: DNA
;   ORGANISM: Glycine max
;   OTHER INFORMATION: Clone ID: 701010921H1
US-09-244-000A-76417


  Query Match          78.2%;  Score 17.2;  DB 16;  Length 253;
  Best Local Similarity   86.4%;  Pred. No. 5.8e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcgtcct 22
          |||||||||||| ||||| | ||
Db     84 cagccgacggtggctgccttct 105


RESULT   8
US-09-654-617-106410
; Sequence 106410, Application US/09654617
; GENERAL INFORMATION:
;  APPLICANT:       Kovalic, David K.
;  APPLICANT:  Liu, Jingdong
;  TITLE OF INVENTION:       Annotated Plant Genes
;  FILE REFERENCE:       38-21(15097)D
;  CURRENT APPLICATION NUMBER: US/09/654,617
;  CURRENT FILING DATE:  2000-09-05
;  NUMBER OF SEQ ID NOS:       463173
; SEQ ID NO 106410
;   LENGTH: 264
;   TYPE: DNA
;   ORGANISM: Glycine max
US-09-654-617-106410


  Query Match          78.2%;  Score 17.2;  DB 25;  Length 264;
  Best Local Similarity   86.4%;  Pred. No. 5.8e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcgtcct 22
          |||||||||||| ||||| | ||
Db     85 cagccgacggtggctgccttct 106


RESULT   9
US-09-684-016-106410
; Sequence 106410, Application US/09684016
; GENERAL INFORMATION:
;  APPLICANT:       Kovalic, David K.
;  APPLICANT:  Liu, Jingdong
;  TITLE OF INVENTION:       Annotated Plant Genes
;  FILE REFERENCE:       38-21(15097)D
;  CURRENT APPLICATION NUMBER: US/09/684,016
;  CURRENT FILING DATE:  2000-10-10
;  PRIOR APPLICATION NUMBER:       US 09/654,617
;  PRIOR FILING DATE:       2000-09-05
;  NUMBER OF SEQ ID NOS:       463173
; SEQ ID NO 106410
;   LENGTH: 264
;   TYPE: DNA
;   ORGANISM: Glycine max
US-09-684-016-106410


  Query Match          78.2%;  Score 17.2;  DB 27;  Length 264;
  Best Local Similarity   86.4%;  Pred. No. 5.8e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcgtcct 22
          |||||||||||| ||||| | ||
Db     85 cagccgacggtggctgccttct 106


RESULT  10
US-09-654-617-109675
; Sequence 109675, Application US/09654617
; GENERAL INFORMATION:
;  APPLICANT:       Kovalic, David K.
;  APPLICANT:  Liu, Jingdong
;  TITLE OF INVENTION:       Annotated Plant Genes
;  FILE REFERENCE:       38-21(15097)D
;  CURRENT APPLICATION NUMBER: US/09/654,617
;  CURRENT FILING DATE:  2000-09-05
;  NUMBER OF SEQ ID NOS:       463173
; SEQ ID NO 109675
;   LENGTH: 425
;   TYPE: DNA
;   ORGANISM: Glycine max
US-09-654-617-109675


  Query Match          78.2%;  Score 17.2;  DB 25;  Length 425;
  Best Local Similarity   86.4%;  Pred. No. 5.9e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcgtcct 22
          |||||||||||| ||||| | ||
Db    227 cagccgacggtggctgccttct 248


RESULT  11
US-09-684-016-109675
; Sequence 109675, Application US/09684016
; GENERAL INFORMATION:
;  APPLICANT:       Kovalic, David K.
;  APPLICANT:  Liu, Jingdong
;  TITLE OF INVENTION:       Annotated Plant Genes
;  FILE REFERENCE:       38-21(15097)D
;  CURRENT APPLICATION NUMBER: US/09/684,016
;  CURRENT FILING DATE:  2000-10-10
;  PRIOR APPLICATION NUMBER:       US 09/654,617
;  PRIOR FILING DATE:       2000-09-05
;  NUMBER OF SEQ ID NOS:       463173
; SEQ ID NO 109675
;   LENGTH: 425
;   TYPE: DNA
;   ORGANISM: Glycine max
US-09-684-016-109675


  Query Match          78.2%;  Score 17.2;  DB 27;  Length 425;
  Best Local Similarity   86.4%;  Pred. No. 5.9e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcgtcct 22
          |||||||||||| ||||| | ||
Db    227 cagccgacggtggctgccttct 248


RESULT  12
US-09-270-849B-122012/c
; Sequence 122012, Application US/09270849B
; GENERAL INFORMATION:
;  APPLICANT: Swimmer et al.
;  TITLE OF INVENTION: Insect genome survey devices
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/09/270,849B
;  CURRENT FILING DATE:  1999-03-17
;  NUMBER OF SEQ ID NOS: 195450
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 122012
;   LENGTH: 426
```

```
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-270-849B-122012


  Query Match             78.2%;  Score 17.2;  DB 16;  Length 426;
  Best Local Similarity   86.4%;  Pred. No. 5.9e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcgtcct 22
          |||||| ||||||| | |||||
Db    247 CAGCCGCCGGTAGCAGTGTCCT 226


RESULT  13
US-60-173-464-23049/c
; Sequence 23049, Application US/60173464
; GENERAL INFORMATION:
;  APPLICANT: Li, Peter W.D.
;  TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS,
;  TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING GPCR PROTEINS AND USES
;  TITLE OF INVENTION: THEREOF
;  FILE REFERENCE: CL000173
;  CURRENT APPLICATION NUMBER: US/60/173,464
;  CURRENT FILING DATE: 1999-12-29
;  NUMBER OF SEQ ID NOS: 30269
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 23049
;   LENGTH: 2339
;   TYPE: DNA
;   ORGANISM: Drosophila
US-60-173-464-23049


  Query Match             78.2%;  Score 17.2;  DB 49;  Length 2339;
  Best Local Similarity   86.4%;  Pred. No. 6e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcgtcct 22
          |||| ||| | ||||||||||
Db    654 CAGCTGACCGGAGCTGCGTCCT 633


RESULT  14
US-60-191-637-28973/c
; Sequence 28973, Application US/60191637
; GENERAL INFORMATION:
;  APPLICANT: Venter, J. Craig
;  TITLE OF INVENTION: NUCLEIC ACID DETECTION KITS COMPRISING
;  TITLE OF INVENTION: GENE SEQUENCES EXPRESSED FROM THE DROSOPHILA GENOME, AND
;  TITLE OF INVENTION: USES THEREOF
;  FILE REFERENCE: CL000392
;  CURRENT APPLICATION NUMBER: US/60/191,637
;  CURRENT FILING DATE: 2000-03-23
;  NUMBER OF SEQ ID NOS: 42660
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 28973
;   LENGTH: 2339
;   TYPE: DNA
;   ORGANISM: DROSOPHILA
US-60-191-637-28973


  Query Match             78.2%;  Score 17.2;  DB 51;  Length 2339;
  Best Local Similarity   86.4%;  Pred. No. 6e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcgtcct 22
          |||| ||| | ||||||||||
Db    654 CAGCTGACCGGAGCTGCGTCCT 633


RESULT  15
US-60-191-681-23456/c
; Sequence 23456, Application US/60191681
; GENERAL INFORMATION:
;  APPLICANT:  Li, Peter, W.D.
;  TITLE OF INVENTION:  ISOLATED DROSOPHILA PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING ISOLATED DROSOPHILA PROTEINS AND
;  TITLE OF INVENTION:  USES THEREOF.
;  FILE REFERENCE:  cl000390
;  CURRENT APPLICATION NUMBER: US/60/191,681
;  CURRENT FILING DATE:  2000-03-23
;  NUMBER OF SEQ ID NOS:  30973
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 23456
;   LENGTH: 2339
;   TYPE: DNA
;   ORGANISM: DROSOPHILA
US-60-191-681-23456


  Query Match             78.2%;  Score 17.2;  DB 51;  Length 2339;
  Best Local Similarity   86.4%;  Pred. No. 6e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcgtcct 22
          |||| ||| | ||||||||||
Db    654 CAGCTGACCGGAGCTGCGTCCT 633


Search completed: September 21, 2001, 02:19:54
Job time: 30466 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: September 21, 2001, 02:39:19 ; Search time 1035.34 Seconds
(without alignments)
42.408 Million cell updates/sec

Title: US-09-138-735-9
Perfect score: 22
Sequence: 1 cagccgacggtagctgcgtcct 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1741792 seqs, 997871973 residues

Total number of hits satisfying chosen parameters: 3483584

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_New:*
1: /cgn1_7/ptodata/2/pna/PCT_NEW_COMB.seq:*
2: /cgn1_7/ptodata/2/pna/US06_NEW_COMB.seq:*
3: /cgn1_7/ptodata/2/pna/US07_NEW_COMB.seq:*
4: /cgn1_7/ptodata/2/pna/US08_NEW_COMB.seq:*
5: /cgn1_7/ptodata/2/pna/US09_NEW_COMB.seq:*
6: /cgn1_7/ptodata/2/pna/US09_NEW_COMB.seq1:*
7: /cgn1_7/ptodata/2/pna/US09_NEW_COMB.seq2:*
8: /cgn1_7/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	22	100.0	22	6	US-09-138-735-9	Sequence 9, Appli
	2	22	100.0	3402	6	US-09-138-735-1	Sequence 1, Appli
	3	17.2	78.2	27082	7	US-09-764-905-25259	Sequence 25259, A
c	4	16.4	74.5	107600	6	US-09-803-736-571	Sequence 571, App
c	5	16.2	73.6	194	8	US-60-252-833-32177	Sequence 32177, A
c	6	16.2	73.6	226	8	US-60-252-833-25761	Sequence 25761, A
c	7	16.2	73.6	264	7	US-09-864-761-24156	Sequence 24156, A
c	8	16.2	73.6	272	8	US-60-252-833-25819	Sequence 25819, A
c	9	16.2	73.6	273	7	US-09-912-292-55206	Sequence 55206, A
c	10	16.2	73.6	277	8	US-60-253-653-2481	Sequence 2481, Ap
c	11	16.2	73.6	326	8	US-60-252-833-32199	Sequence 32199, A
c	12	16.2	73.6	327	8	US-60-253-652-13183	Sequence 13183, A
	13	16.2	73.6	332	7	US-09-540-213-15508	Sequence 15508, A
c	14	16.2	73.6	345	8	US-60-253-652-14207	Sequence 14207, A
c	15	16.2	73.6	369	8	US-60-252-833-19140	Sequence 19140, A
c	16	16.2	73.6	378	6	US-09-909-627-8996	Sequence 8996, Ap
c	17	16.2	73.6	380	8	US-60-252-833-32484	Sequence 32484, A
c	18	16.2	73.6	412	8	US-60-252-833-32592	Sequence 32592, A
c	19	16.2	73.6	412	8	US-60-252-833-32600	Sequence 32600, A
c	20	16.2	73.6	421	8	US-60-253-651-22049	Sequence 22049, A
	21	16.2	73.6	471	6	US-09-866-555-17442	Sequence 17442, A
c	22	16.2	73.6	482	8	US-60-253-653-3881	Sequence 3881, Ap
c	23	16.2	73.6	491	6	US-09-849-626-1155	Sequence 1155, Ap
c	24	16.2	73.6	491	7	US-09-902-941-1155	Sequence 1155, Ap
	25	16.2	73.6	493	7	US-09-834-366-50093	Sequence 50093, A
	26	16.2	73.6	520	6	US-09-849-626-927	Sequence 927, App
	27	16.2	73.6	520	7	US-09-902-941-927	Sequence 927, App
c	28	16.2	73.6	540	8	US-60-253-653-15892	Sequence 15892, A
c	29	16.2	73.6	542	8	US-60-252-833-30020	Sequence 30020, A
c	30	16.2	73.6	546	8	US-60-253-652-16860	Sequence 16860, A
c	31	16.2	73.6	593	7	US-09-864-761-7514	Sequence 7514, Ap
	32	16.2	73.6	613	8	US-60-253-654-10376	Sequence 10376, A
	33	16.2	73.6	613	8	US-60-255-592-10376	Sequence 10376, A
c	34	16.2	73.6	725	8	US-60-252-833-9958	Sequence 9958, Ap
c	35	16.2	73.6	731	8	US-60-253-652-14151	Sequence 14151, A
c	36	16.2	73.6	740	8	US-60-252-833-27687	Sequence 27687, A
	37	16.2	73.6	1029	8	US-60-312-544-1195	Sequence 1195, Ap
	38	16.2	73.6	5642	8	US-60-278-561-9583	Sequence 9583, Ap
c	39	16.2	73.6	5643	6	US-09-898-888-11040	Sequence 11040, A
c	40	15.8	71.8	298	8	US-60-253-654-18217	Sequence 18217, A
c	41	15.8	71.8	298	8	US-60-255-592-18217	Sequence 18217, A
	42	15.8	71.8	383	8	US-60-253-456-3926	Sequence 3926, Ap
c	43	15.8	71.8	404	6	US-09-849-626-443	Sequence 443, App
c	44	15.8	71.8	404	7	US-09-902-941-443	Sequence 443, App
	45	15.8	71.8	5031	6	US-09-764-877-2950	Sequence 2950, Ap

ALIGNMENTS

RESULT 1
US-09-138-735-9
; Sequence 9, Application US/09138735
; GENERAL INFORMATION:
; APPLICANT: PARANHOS-BACCALA, Glaucia
; APPLICANT: LESENECHAL, Mylene
; APPLICANT: JOLIVET, Michel
; TITLE OF INVENTION: TRYPANOSOMA CRUZI ANTIGEN, GENE ENCODING THEREFOR AND METHODS
; TITLE OF INVENTION: DETECTING AND TREATING CHAGAS DISEASE
; FILE REFERENCE: WPB 36400B
; CURRENT APPLICATION NUMBER: US/09/138,735
; CURRENT FILING DATE: 1998-08-24
; PRIOR APPLICATION NUMBER: US 08/480,917
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: FR 94/10132
; PRIOR FILING DATE: 1994-08-12
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 22
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: probe/primer
US-09-138-735-9

Query Match 100.0%; Score 22; DB 6; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.26;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 cagccgacggtagctgcgtcct 22
          ||||||||||||||||||||||
Db      1 cagccgacggtagctgcgtcct 22
```

RESULT 2
US-09-138-735-1
; Sequence 1, Application US/09138735
; GENERAL INFORMATION:
; APPLICANT: PARANHOS-BACCALA, Glaucia
; APPLICANT: LESENECHAL, Mylene
; APPLICANT: JOLIVET, Michel
; TITLE OF INVENTION: TRYPANOSOMA CRUZI ANTIGEN, GENE ENCODING THEREFOR AND METHODS
; TITLE OF INVENTION: DETECTING AND TREATING CHAGAS DISEASE
; FILE REFERENCE: WPB 36400B
; CURRENT APPLICATION NUMBER: US/09/138,735
; CURRENT FILING DATE: 1998-08-24

; PRIOR APPLICATION NUMBER: US 08/480,917
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: FR 94/10132
; PRIOR FILING DATE: 1994-08-12
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 3402
; TYPE: DNA
; ORGANISM: Trypanosoma cruzi
US-09-138-735-1

Query Match 100.0%; Score 22; DB 6; Length 3402;
Best Local Similarity 100.0%; Pred. No. 0.29;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cagccgacggtagctgcgtcct 22
||||||||||||||||||||||
Db 1266 cagccgacggtagctgcgtcct 1287

RESULT 3
US-09-764-905-25259
; Sequence 25259, Application US/09764905
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC004
; CURRENT APPLICATION NUMBER: US/09/764,905
; CURRENT FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/225,757
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/226,868
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/216,647
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/225,267
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/216,880
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/225,270
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/251,869
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/235,834
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/234,274
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: 60/234,223
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: 60/228,924
; PRIOR FILING DATE: 2000-08-30
; PRIOR APPLICATION NUMBER: 60/224,518
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/236,369
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/224,519
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,964
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/241,809
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/249,299
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/236,327
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/241,785
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/244,617
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 60/225,268
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/236,368
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/251,856
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/251,868
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/229,344
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/234,997
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: 60/229,343
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,345
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,287
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,513
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: 60/231,413
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/229,509
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: 60/236,367
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/237,039
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/237,038
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/236,370
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/236,802
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/237,037
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/237,040
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/240,960
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/239,935
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 60/239,937
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 60/241,787
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/246,474
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/246,532
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/249,216
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,210
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/226,681
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/225,759

; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/225,213
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/227,182
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/225,214
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/235,836
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/230,438
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/215,135
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 60/225,266
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/249,218
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,208
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,213
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,212
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,207
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,245
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,244
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,217
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,211
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,215
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,264
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,214
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,297
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/232,400
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/231,242
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,081
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,080
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,414
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,244
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/233,064
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/233,063
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,397
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,399
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,401
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/241,808
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,826
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,786
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,221
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/246,475
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/231,243
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/233,065
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,398

```
  Query Match             78.2%;  Score 17.2;  DB 7;  Length 27082;
  Best Local Similarity   86.4%;  Pred. No. 63;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 cagccgacggtagctgcgtcct 22
           ||||| |||||||||| | ||
Db     531 cagcccacggtagctgctttct 552
```

RESULT 4
US-09-803-736-571/c
; Sequence 571, Application US/09803736
; GENERAL INFORMATION:
; APPLICANT: Bush, David F.
; APPLICANT: Levin, Irena M.
; APPLICANT: Norris, Susan R.
; APPLICANT: Rounsley, Steven D.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Plant Polymorphic Markers and Uses Thereof
; FILE REFERENCE: 38-10(15493)D
; CURRENT APPLICATION NUMBER: US/09/803,736
; CURRENT FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/534,859
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: identified by Attorney Docket number 04983.0206CPUS01 38
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 1582
; SEQ ID NO 571
; LENGTH: 107600
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-803-736-571

```
  Query Match             74.5%;  Score 16.4;  DB 6;  Length 107600;
  Best Local Similarity   94.4%;  Pred. No. 1.6e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5 cgacggtagctgcgtcct 22
           ||||||||||||| ||||
Db   33860 CGACGGTAGCTGCTTCCT 33843
```

RESULT 5
US-60-252-833-32177/c
; Sequence 32177, Application US/60252833
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Grigor, Murray R
; TITLE OF INVENTION: Compositions isolated from bovine
; TITLE OF INVENTION: tissues and methods for their use.
; FILE REFERENCE: 1052P2
; CURRENT APPLICATION NUMBER: US/60/252,833
; CURRENT FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 43535
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 32177
; LENGTH: 194
; TYPE: DNA
; ORGANISM: Bovine
US-60-252-833-32177

```
  Query Match             73.6%;  Score 16.2;  DB 8;  Length 194;
  Best Local Similarity   85.7%;  Pred. No. 1.7e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;
```

```
Qy      2 agccgacggtagctgcgtcct 22
          |||| ||| | ||||||||||
Db    179 AGCCCACGATGGCTGCGTCCT 159


RESULT   6
US-60-252-833-25761/c
; Sequence 25761, Application US/60252833
; GENERAL INFORMATION:
;  APPLICANT: Glenn, Matthew
;  APPLICANT:  Grigor, Murray R
;  TITLE OF INVENTION: Compositions isolated from bovine
;  TITLE OF INVENTION:  tissues and methods for their use.
;  FILE REFERENCE: 1052P2
;  CURRENT APPLICATION NUMBER: US/60/252,833
;  CURRENT FILING DATE:  2000-11-21
;  NUMBER OF SEQ ID NOS: 43535
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25761
;   LENGTH: 226
;   TYPE: DNA
;   ORGANISM: Bovine
US-60-252-833-25761


  Query Match             73.6%;  Score 16.2;  DB 8;  Length 226;
  Best Local Similarity   85.7%;  Pred. No. 1.7e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 agccgacggtagctgcgtcct 22
          |||| ||| | ||||||||||
Db    208 AGCCCACGATGGCTGCGTCCT 188


RESULT   7
US-09-864-761-24156/c
; Sequence 24156, Application US/09864761
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 24156
;   LENGTH: 264
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL121897.15
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.9
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.4
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.4
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.5
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.5
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
;   OTHER INFORMATION: SWISSPROT HIT: O21403, EVALUE 5.50e+00
;   OTHER INFORMATION: EST_HUMAN HIT: BE900549.1, EVALUE 0.00e+00
;   OTHER INFORMATION: NT HIT: D83784.1, EVALUE 0.00e+00
US-09-864-761-24156


  Query Match             73.6%;  Score 16.2;  DB 7;  Length 264;
  Best Local Similarity   85.7%;  Pred. No. 1.7e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcgtcc 21
          ||||||| ||| ||||||| ||
Db    157 CAGCCGGCGGCAGCTGCGGCC 137


RESULT   8
US-60-252-833-25819/c
; Sequence 25819, Application US/60252833
; GENERAL INFORMATION:
;  APPLICANT: Glenn, Matthew
;  APPLICANT:  Grigor, Murray R
;  TITLE OF INVENTION: Compositions isolated from bovine
;  TITLE OF INVENTION:  tissues and methods for their use.
;  FILE REFERENCE: 1052P2
;  CURRENT APPLICATION NUMBER: US/60/252,833
;  CURRENT FILING DATE:  2000-11-21
;  NUMBER OF SEQ ID NOS: 43535
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25819
;   LENGTH: 272
;   TYPE: DNA
;   ORGANISM: Bovine
US-60-252-833-25819


  Query Match             73.6%;  Score 16.2;  DB 8;  Length 272;
  Best Local Similarity   85.7%;  Pred. No. 1.7e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 agccgacggtagctgcgtcct 22
          |||| ||| | ||||||||||
Db    208 AGCCCACGATGGCTGCGTCCT 188


RESULT   9
US-09-912-292-55206/c
; Sequence 55206, Application US/09912292
```

; GENERAL INFORMATION:
; APPLICANT: Rosen, et. al.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 101
; FILE REFERENCE: PO-101
; CURRENT APPLICATION NUMBER: US/09/912,292
; CURRENT FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: 08/103,746
; PRIOR FILING DATE: 1993-08-09
; PRIOR APPLICATION NUMBER: 09/859,417
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 08/103,743
; PRIOR FILING DATE: 1993-08-09
; PRIOR APPLICATION NUMBER: 09/741,827
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 09/813,154
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 08/196,482
; PRIOR FILING DATE: 1994-02-15
; PRIOR APPLICATION NUMBER: 09/783,587
; PRIOR FILING DATE: 2001-02-15
; PRIOR APPLICATION NUMBER: 08/196,481
; PRIOR FILING DATE: 1994-02-15
; PRIOR APPLICATION NUMBER: 08/220,662
; PRIOR FILING DATE: 1994-03-31
; PRIOR APPLICATION NUMBER: 08/220,661
; PRIOR FILING DATE: 1994-03-31
; PRIOR APPLICATION NUMBER: 08/275,627
; PRIOR FILING DATE: 1994-07-15
; PRIOR APPLICATION NUMBER: 08/345,704
; PRIOR FILING DATE: 1994-11-21
; PRIOR APPLICATION NUMBER: 09/859,662
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 08/401,881
; PRIOR FILING DATE: 1995-03-10
; PRIOR APPLICATION NUMBER: 09/782,161
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: 08/790,776
; PRIOR FILING DATE: 1997-01-30
; PRIOR APPLICATION NUMBER: 60/010,803
; PRIOR FILING DATE: 1996-01-30
; PRIOR APPLICATION NUMBER: 08/799,180
; PRIOR FILING DATE: 1997-02-12
; PRIOR APPLICATION NUMBER: 60/011,618
; PRIOR FILING DATE: 1996-02-13
; PRIOR APPLICATION NUMBER: 08/803,609
; PRIOR FILING DATE: 1997-02-21
; PRIOR APPLICATION NUMBER: 60/011,985
; PRIOR FILING DATE: 1996-02-21
; PRIOR APPLICATION NUMBER: 08/971,050
; PRIOR FILING DATE: 1997-11-14
; PRIOR APPLICATION NUMBER: 60/031,038
; PRIOR FILING DATE: 1996-11-18
; PRIOR APPLICATION NUMBER: 08/975,985
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 60/031,221
; PRIOR FILING DATE: 1996-11-25
; PRIOR APPLICATION NUMBER: 08/985,366
; PRIOR FILING DATE: 1997-12-04
; PRIOR APPLICATION NUMBER: 09/842,827
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: 60/032,782
; PRIOR FILING DATE: 1996-12-06
; PRIOR APPLICATION NUMBER: 09/076,898
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/047,004
; PRIOR FILING DATE: 1997-05-13
; PRIOR APPLICATION NUMBER: 09/078,614
; PRIOR FILING DATE: 1998-05-14
; PRIOR APPLICATION NUMBER: 60/046,498
; PRIOR FILING DATE: 1997-05-14
; PRIOR APPLICATION NUMBER: 09/189,834
; PRIOR FILING DATE: 1998-11-12
; PRIOR APPLICATION NUMBER: 60/065,454
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 09/366,690
; PRIOR FILING DATE: 1999-08-04
; PRIOR APPLICATION NUMBER: 60/095,484
; PRIOR FILING DATE: 1998-08-05
; PRIOR APPLICATION NUMBER: 09/342,216
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: 60/091,171
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: 09/342,061
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: 60/091,170
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: 09/487,572
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: 60/116,678
; PRIOR FILING DATE: 1999-01-21
; PRIOR APPLICATION NUMBER: 09/497,772
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/118,784
; PRIOR FILING DATE: 1999-02-05
; PRIOR APPLICATION NUMBER: 09/504,577
; PRIOR FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 60/120,434
; PRIOR FILING DATE: 1999-02-16
; PRIOR APPLICATION NUMBER: 09/517,011
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/122,447
; PRIOR FILING DATE: 1999-03-02
; PRIOR APPLICATION NUMBER: 09/628,857
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: 60/145,956
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: 09/628,858
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: 60/145,955
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: 09/741,033
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 09/773,518
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/171,621
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 09/796,765
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: 09/867,682
; PRIOR FILING DATE: 2001-05-31
; PRIOR APPLICATION NUMBER: 60/187,016
; PRIOR FILING DATE: 2000-03-06
; PRIOR APPLICATION NUMBER: 60/225,269
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/225,629
; PRIOR FILING DATE: 2000-08-15
; NUMBER OF SEQ ID NOS: 55553
; SEQ ID NO 55206
; LENGTH: 273
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (42)..(42)
; OTHER INFORMATION: n is equal to a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (56)..(56)
; OTHER INFORMATION: n is equal to a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (69)..(69)
; OTHER INFORMATION: n is equal to a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (125)..(126)
; OTHER INFORMATION: n is equal to a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (176)..(176)

```
;   OTHER INFORMATION: n is equal to a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (199)..(199)
;   OTHER INFORMATION: n is equal to a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (244)..(244)
;   OTHER INFORMATION: n is equal to a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (247)..(247)
;   OTHER INFORMATION: n is equal to a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (260)..(260)
;   OTHER INFORMATION: n is equal to a,t,g, or c
US-09-912-292-55206


  Query Match            73.6%;  Score 16.2;  DB 7;  Length 273;
  Best Local Similarity  85.7%;  Pred. No. 1.7e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 agccgacggtagctgcgtcct 22
          |||| |||||||||||  || |
Db    236 AGCCTACGGTAGCTGCTTCTT 216


RESULT  10
US-60-253-653-2481/c
; Sequence 2481, Application US/60253653
; GENERAL INFORMATION:
;  APPLICANT: Glenn, Matthew
;  TITLE OF INVENTION: Compositions isolated from ovine tissues
;  TITLE OF INVENTION:  and methods for their use.
;  FILE REFERENCE: 1057P
;  CURRENT APPLICATION NUMBER: US/60/253,653
;  CURRENT FILING DATE:  2000-11-28
;  NUMBER OF SEQ ID NOS: 30124
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2481
;   LENGTH: 277
;   TYPE: DNA
;   ORGANISM: Ovine
US-60-253-653-2481


  Query Match            73.6%;  Score 16.2;  DB 8;  Length 277;
  Best Local Similarity  85.7%;  Pred. No. 1.7e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 agccgacggtagctgcgtcct 22
          |||| ||| | ||||||||||
Db    211 AGCCCACGATGGCTGCGTCCT 191


RESULT  11
US-60-252-833-32199/c
; Sequence 32199, Application US/60252833
; GENERAL INFORMATION:
;  APPLICANT: Glenn, Matthew
;  APPLICANT:  Grigor, Murray R
;  TITLE OF INVENTION: Compositions isolated from bovine
;  TITLE OF INVENTION:  tissues and methods for their use.
;  FILE REFERENCE: 1052P2
;  CURRENT APPLICATION NUMBER: US/60/252,833
;  CURRENT FILING DATE:  2000-11-21
;  NUMBER OF SEQ ID NOS: 43535
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 32199
;   LENGTH: 326
;   TYPE: DNA
;   ORGANISM: Bovine
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(326)
;   OTHER INFORMATION: n = A,T,C or G
US-60-252-833-32199


  Query Match            73.6%;  Score 16.2;  DB 8;  Length 326;
  Best Local Similarity  85.7%;  Pred. No. 1.7e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 agccgacggtagctgcgtcct 22
          |||| ||| | ||||||||||
Db    210 AGCCCACGATGGCTGCGTCCT 190


RESULT  12
US-60-253-652-13183/c
; Sequence 13183, Application US/60253652
; GENERAL INFORMATION:
;  APPLICANT: Glenn, Matthew
;  APPLICANT:  Grigor, Murray R
;  TITLE OF INVENTION: Compositions isolated from bovine
;  TITLE OF INVENTION:  tissues and methods for their use.
;  FILE REFERENCE: 1055P2
;  CURRENT APPLICATION NUMBER: US/60/253,652
;  CURRENT FILING DATE:  2000-11-28
;  NUMBER OF SEQ ID NOS: 29954
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13183
;   LENGTH: 327
;   TYPE: DNA
;   ORGANISM: Bovine
US-60-253-652-13183


  Query Match            73.6%;  Score 16.2;  DB 8;  Length 327;
  Best Local Similarity  85.7%;  Pred. No. 1.7e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 agccgacggtagctgcgtcct 22
          |||| ||| | ||||||||||
Db    215 AGCCCACGATGGCTGCGTCCT 195


RESULT  13
US-09-540-213-15508
; Sequence 15508, Application US/09540213
; GENERAL INFORMATION:
;  APPLICANT: Seilhamer, Jeffrey J.
;  APPLICANT:  Delegeane, Angelo M.
;  APPLICANT:  Stuart, Susan G.
;  APPLICANT:  Stuve, Laura L.
;  APPLICANT:  Mullahy, Sara J.
;  APPLICANT:  Naughton, Rebecca E.
;  TITLE OF INVENTION: POLYNUCLEOTIDES OF LIVER AND PANCREAS
;  FILE REFERENCE: PD-1031 CIP
;  CURRENT APPLICATION NUMBER: US/09/540,213
;  CURRENT FILING DATE:  2000-03-31
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 53844
; SEQ ID NO 15508
;   LENGTH: 332
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: hu01083786
US-09-540-213-15508


  Query Match            73.6%;  Score 16.2;  DB 7;  Length 332;
  Best Local Similarity  85.7%;  Pred. No. 1.7e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 cagccgacggtagctgcgtcc 21
```

```
          ||||||  || ||||||||||
Db     85 cagccggaggcagctgcgtcc 105


RESULT  14
US-60-253-652-14207/c
; Sequence 14207, Application US/60253652
; GENERAL INFORMATION:
;  APPLICANT: Glenn, Matthew
;  APPLICANT:  Grigor, Murray R
;  TITLE OF INVENTION: Compositions isolated from bovine
;  TITLE OF INVENTION:  tissues and methods for their use.
;  FILE REFERENCE: 1055P2
;  CURRENT APPLICATION NUMBER: US/60/253,652
;  CURRENT FILING DATE:  2000-11-28
;  NUMBER OF SEQ ID NOS: 29954
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14207
;   LENGTH: 345
;   TYPE: DNA
;   ORGANISM: Bovine
US-60-253-652-14207


  Query Match             73.6%;  Score 16.2;  DB 8;  Length 345;
  Best Local Similarity   85.7%;  Pred. No. 1.7e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 agccgacggtagctgcgtcct 22
           |||| ||| | ||||||||||
Db     191 AGCCCACGATGGCTGCGTCCT 171


RESULT  15
US-60-252-833-19140/c
; Sequence 19140, Application US/60252833
; GENERAL INFORMATION:
;  APPLICANT: Glenn, Matthew
;  APPLICANT:  Grigor, Murray R
;  TITLE OF INVENTION: Compositions isolated from bovine
;  TITLE OF INVENTION:  tissues and methods for their use.
;  FILE REFERENCE: 1052P2
;  CURRENT APPLICATION NUMBER: US/60/252,833
;  CURRENT FILING DATE:  2000-11-21
;  NUMBER OF SEQ ID NOS: 43535
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 19140
;   LENGTH: 369
;   TYPE: DNA
;   ORGANISM: Bovine
US-60-252-833-19140


  Query Match             73.6%;  Score 16.2;  DB 8;  Length 369;
  Best Local Similarity   85.7%;  Pred. No. 1.8e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 agccgacggtagctgcgtcct 22
           |||| ||| | ||||||||||
Db     208 AGCCCACGATGGCTGCGTCCT 188


Search completed: September 21, 2001, 02:39:25
Job time: 15051 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: September 21, 2001, 02:42:32 ; Search time 330.77 Seconds
(without alignments)
34.169 Million cell updates/sec

Title: US-09-138-735-8
Perfect score: 18
Sequence: 1 tgcagcagcggcagaagt 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0601:*
1: /cgn1_9/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /cgn1_9/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /cgn1_9/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /cgn1_9/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /cgn1_9/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /cgn1_9/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /cgn1_9/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /cgn1_9/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /cgn1_9/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /cgn1_9/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /cgn1_9/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /cgn1_9/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /cgn1_9/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /cgn1_9/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /cgn1_9/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /cgn1_9/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /cgn1_9/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /cgn1_9/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	18	100.0	18	17	AAT27312	T.cruzi epimastigo
	2	18	100.0	18	20	AAX84094	PCR primer for T.
c	3	18	100.0	3402	17	AAT27310	Trypanosoma cruzi
c	4	18	100.0	3402	20	AAX84092	T. cruzi PTc40 cod
	5	18	100.0	11878	19	AAV30199	Protein kinase cat
	6	18	100.0	11883	19	AAV30198	Protein kinase cat
	7	16.4	91.1	876	8	AAN70094	Sequence of region
	8	16.4	91.1	36519	19	AAV22141	Chimpanzee adenovi
c	9	15.4	85.6	20	21	AAA14833	PCR primer HG52.16
	10	15.4	85.6	39	20	AAX06966	Bacillus agaradhae
	11	15.4	85.6	240	21	AAA35346	Myrtaceae microsat
	12	15.4	85.6	303	16	AAT18914	DNA encoding spide
c	13	15.4	85.6	328	21	AAC12718	Human secreted pro
	14	15.4	85.6	341	21	AAC03564	Human secreted pro
	15	15.4	85.6	422	21	AAC79295	Human lung tumour-
	16	15.4	85.6	448	21	AAA95498	Human p60 anti-vir
	17	15.4	85.6	457	21	AAA89684	Mouse Exo92 nucleo
	18	15.4	85.6	478	21	AAA89657	Exo71 nucleotide s
	19	15.4	85.6	502	21	AAA89683	Mouse Exo91 nucleo
c	20	15.4	85.6	1269	21	AAC42299	Arabidopsis thalia
	21	15.4	85.6	1276	21	AAA15086	Transcript SCP-3B
	22	15.4	85.6	1381	12	AAQ14472	E1-beta subunit of
	23	15.4	85.6	1558	21	AAC49642	Arabidopsis thalia
	24	15.4	85.6	1837	21	AAC77971	Human cancer assoc
c	25	15.4	85.6	1955	21	AAZ99588	cDNA encoding a hu
	26	15.4	85.6	2000	16	AAQ87925	Erythroid p55 geno
	27	15.4	85.6	2030	21	AAC98843	Human pancreatic c
	28	15.4	85.6	2177	21	AAZ57853	Protein regulating
	29	15.4	85.6	2204	21	AAC77452	Human ORFX ORF3007
c	30	15.4	85.6	2214	19	AAV22682	New DNA sequence i
	31	15.4	85.6	3000	21	AAA99899	Aspergillus nidula
	32	15.4	85.6	3183	20	AAX26611	Trypanosoma cruzi
c	33	15.4	85.6	3212	22	AAC84373	Nucleotide sequenc
c	34	15.4	85.6	3331	19	AAV22683	New DNA sequence i
c	35	15.4	85.6	3331	20	AAZ32021	Human METH1 relate
c	36	15.4	85.6	3331	22	AAC90078	D86074 cDNA clone.
	37	15.4	85.6	3426	20	AAZ23819	Human SREBP-2 cDNA
	38	15.4	85.6	5197	16	AAQ79038	Human SREBP-2 codi
	39	15.4	85.6	5228	20	AAV81384	Nucleic acid encod
c	40	15.4	85.6	6115	21	AAZ46519	Murine CACNAIF enc
	41	15.4	85.6	34094	20	AAZ30163	Complete nucleotid
	42	15.4	85.6	273254	21	AAC81914	Chlamydia pneumoni
	43	15	83.3	362	22	AAF93718	cDNA encoding SRT
	44	15	83.3	426	20	AAX08739	Nucleotide encodin
c	45	15	83.3	1065	21	AAF14557	Aspergillus oryzae

ALIGNMENTS

RESULT 1
AAT27312
ID AAT27312 standard; cDNA; 18 BP.
XX
AC AAT27312;
XX
DT 26-NOV-1996 (first entry)
XX
DE T.cruzi epimastigotic PTc100t antigen primer corresp. to bases 1442-59.
XX
KW Antigen; Trypanosoma cruzi; epimastigote; serum; Chagas disease; probe;
KW Primer; PCR; polymerase chain reaction; amplification; antibody; ss.
XX
OS Synthetic.
XX
PN FR2723589-A1.
XX
PD 16-FEB-1996.
XX
PF 12-AUG-1994; 94FR-0010132.
XX
PR 12-AUG-1994; 94FR-0010132.
XX
PA (INMR) BIO MERIEUX.
XX
PI Jolivet M, Lesenechal M, Paranhos-Baccala G;
XX
DR WPI; 1996-190287/20.
XX
PT New nucleic acid encoding Trypanosoma cruzi epimastigotic antigen -
PT useful for diagnosis, monitoring and therapy of Chagas disease
XX
PS Claim 29; Page 36; 55pp; French.
XX

```
CC   The primers AAT27311-5 were used to PCR amplify the sequence encoding a
CC   novel isolated antigenic protein from Trypanosoma cruzi epimastigotes,
CC   designated PTc100t (AAT27310). The primers, derived from the sequences
CC   of a 594 and 1041 bp fragment of PTc100t, amplified the gene as 3
CC   fragments. This primer corresponds to nucleotides 1442-59 of the PT100t
CC   sequence, derived from the 594 bp fragment. The 594 bp fragment was
CC   isolated from a T.cruzi genomic expression library in lambda gt11, using
CC   a mixture of sera from patients with Chagas disease. It corresponds to
CC   nucleotides 1232-1825 of PTc100t. The 1041 bp fragment was isolated from
CC   a lambda gt10 library using the 594 bp fragment as a probe. The protein,
CC   or antibodies raised against it, can be used in the detection and
CC   monitoring of T.cruzi infection i.e. Chagas disease.
XX
SQ   Sequence 18 BP; 5 A; 4 C; 7 G; 2 T; 0 other;

  Query Match             100.0%;  Score 18;  DB 17;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||||||||||||||||
Db      1 tgcagcagcggcagaagt 18


RESULT   2
AAX84094
ID   AAX84094 standard; DNA; 18 BP.
XX
AC   AAX84094;
XX
DT   27-AUG-1999  (first entry)
XX
DE   PCR primer for T. cruzi PTc40 coding sequence.
XX
KW   PTc40; Tc40; infection; diagnosis; immune complex; antigenic determinant;
KW   therapy; antibody; PCR primer;  ss.
XX
OS   Synthetic.
OS   Trypanosoma cruzi.
XX
PN   WO9929867-A1.
XX
PD   17-JUN-1999.
XX
PF   10-DEC-1998;   98WO-IB01987.
XX
PR   10-DEC-1997;   97US-0988242.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Jolivet M,  Lesenechal M,  Mandrand B,  Paranhos-Baccala G;
XX
DR   WPI; 1999-394978/33.
XX
PT   New Trypanosoma cruzi antigen
XX
PS   Disclosure; Page 21; 65pp; English.
XX
CC   This sequence is a PCR primer for DNA encoding the Trypanosoma cruzi
CC   PTc40 protein of the invention. The PTc40 antigenic determinant is
CC   useful as a reagent for detection and/or monitoring of Trypanosoma cruzi
CC   infection from samples including blood serum or plasma, urine, saliva, or
CC   tears, by contacting with the sample and detecting an immune complex. The
CC   PTc40 antigenic determinant, the vector, expression cassette, cell or
CC   antibody are useful for treatment or prevention (vaccine) of a
CC   Trypanosoma cruzi infection in a man or animal. Current Trypanosoma cruzi
CC   antigens are obtained from protein fractions of the noninfectious stage
CC   of the parasite, and these do not allow sufficient production of antigens
CC   for use in reliable serological diagnostic tests. The strain to strain
CC   polymorphism reduces reliability of the tests.
XX
SQ   Sequence 18 BP; 5 A; 4 C; 7 G; 2 T; 0 other;

  Query Match             100.0%;  Score 18;  DB 20;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||||||||||||||||
Db      1 tgcagcagcggcagaagt 18


RESULT   3
AAT27310/c
ID   AAT27310 standard; cDNA; 3402 BP.
XX
AC   AAT27310;
XX
DT   26-NOV-1996  (first entry)
XX
DE   Trypanosoma cruzi epimastigotic PTc100t antigen gene.
XX
KW   Antigen; Trypanosoma cruzi; epimastigote; serum; Chagas disease; probe;
KW   Primer; PCR; polymerase chain reaction; amplification; antibody; ds.
XX
OS   Trypanosoma cruzi.
XX
FH   Key             Location/Qualifiers
FT   CDS             266..3013
FT                   /*tag= a
FT                   /product= PTc100t epimastogotic antigen
XX
PN   FR2723589-A1.
XX
PD   16-FEB-1996.
XX
PF   12-AUG-1994;   94FR-0010132.
XX
PR   12-AUG-1994;   94FR-0010132.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Jolivet M,  Lesenechal M,  Paranhos-Baccala G;
XX
DR   WPI; 1996-190287/20.
DR   P-PSDB; AAR91615.
XX
PT   New nucleic acid encoding Trypanosoma cruzi epimastigotic antigen -
PT   useful for diagnosis, monitoring and therapy of Chagas disease
XX
PS   Claim 1; Page 24-26; 55pp; French.
XX
CC   This is the nucleotide sequence encoding a novel isolated antigenic
CC   protein from Trypanosoma cruzi epimastigotes, designated PTc100t.
CC   The clone Tc50 was isolated from a T.cruzi genomic expression library in
CC   lambda gt11, using a mixture of sera from patients with Chagas disease.
CC   Clone Tc50 contained an 594 bp insert corresp. to nucleotides 1232-1825
CC   of this sequence.  The Tc50 sequence was subsequently used to probe a
CC   Southern blot of restriction enzyme digested T.cruzi DNA and also screen
CC   a lambda gt10 library to isolate a 1041 bp EcoRI fragment corresp. to
CC   nucleotides 1403-2443 of PTc100t.  Primers (AAT27311-5) were synthesised
CC   based on the sequences of the 594 and 1041 bp fragments and used to
CC   amplify the PTc100t clone as 3 fragments from cDNA derived from mRNA
CC   purified from T.cruzi epimastigotes.  The protein or antibodies raised
CC   against it can be used in the detection and monitoring of T.cruzi
CC   infection i.e. Chagas disease.
XX
SQ   Sequence 3402 BP; 889 A; 818 C; 958 G; 737 T; 0 other;

  Query Match             100.0%;  Score 18;  DB 17;  Length 3402;
  Best Local Similarity   100.0%;  Pred. No. 43;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      1 tgcagcagcggcagaagt 18
          ||||||||||||||||||
Db   1459 TGCAGCAGCGGCAGAAGT 1442


RESULT   4
AAX84092/c
ID   AAX84092 standard; cDNA; 3402 BP.
XX
AC   AAX84092;
XX
DT   27-AUG-1999  (first entry)
XX
DE   T. cruzi PTc40 coding sequence.
XX
KW   PTc40; Tc40; infection; diagnosis; immune complex; antigenic determinant;
KW   therapy; antibody; ds.
XX
OS   Trypanosoma cruzi.
XX
PN   WO9929867-A1.
XX
PD   17-JUN-1999.
XX
PF   10-DEC-1998;   98WO-IB01987.
XX
PR   10-DEC-1997;   97US-0988242.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Jolivet M,  Lesenechal M,  Mandrand B,  Paranhos-Baccala G;
XX
DR   WPI; 1999-394978/33.
DR   P-PSDB; AAY22124.
XX
PT   New Trypanosoma cruzi antigen
XX
PS   Claim 1; Page 52-56; 65pp; English.
XX
CC   This sequence encodes the Trypanosoma cruzi PTc40 protein of the
CC   invention, and is designated Tc40. The PTc40 antigenic determinant is
CC   useful as a reagent for detection and/or monitoring of Trypanosoma cruzi
CC   infection from samples including blood serum or plasma, urine, saliva, or
CC   tears, by contacting with the sample and detecting an immune complex. The
CC   PTc40 antigenic determinant, the vector, expression cassette, cell or
CC   antibody are useful for treatment or prevention (vaccine) of a
CC   Trypanosoma cruzi infection in a man or animal. Current Trypanosoma cruzi
CC   antigens are obtained from protein fractions of the noninfectious stage
CC   of the parasite, and these do not allow sufficient production of antigens
CC   for use in reliable serological diagnostic tests. The strain to strain
CC   polymorphism reduces reliability of the tests.
XX
SQ   Sequence 3402 BP; 888 A; 821 C; 956 G; 737 T; 0 other;


  Query Match             100.0%;  Score 18;  DB 20;  Length 3402;
  Best Local Similarity   100.0%;  Pred. No. 43;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||||||||||||||||
Db   1459 TGCAGCAGCGGCAGAAGT 1442


RESULT   5
AAV30199
ID   AAV30199 standard; DNA; 11878 BP.
XX
AC   AAV30199;
XX
DT   14-SEP-1998  (first entry)
XX
DE   Protein kinase catalytic subunit gene.
XX
KW   Severe combined immunodeficiency disease; SCID; horse; diagnosis;
KW   DNA-dependent protein kinase; ds.
XX
OS   Equus caballus.
XX
PN   WO9821367-A1.
XX
PD   22-MAY-1998.
XX
PF   14-NOV-1997;   97WO-US21066.
XX
PR   15-NOV-1996;   96US-0031261.
XX
PA   (TEXA ) UNIV TEXAS SYSTEM.
XX
PI   Meeks K;
XX
DR   WPI; 1998-297967/26.
XX
PT   DNA-dependent protein kinase catalytic subunit - useful for
PT   determining equine severe combined immunodeficiency alleles
XX
PS   Disclosure; Page 70-78; 98pp; English.
XX
CC   This isolated DNA molecule encodes an equine DNA-dependent protein
CC   kinase catalytic subunit (DNA-PKcs).  A claimed method of
CC   identifying an Arabian horse that is a carrier of equine severe
CC   combined immunodeficiency (SCID) comprises determining whether the
CC   horse has a mutation in a SCID determinant region of the DNA-PKcs
CC   gene (see also AAV30196 and AAV30197).  Sequence analysis of DNA-PKsc
CC   genes from normal and SCID equine fibroblasts shows that a 5 bp
CC   deletion is present in SCID foals at a site that corresponds to
CC   nucleotide 9454 of the 12,381 nucleotide coding sequence of the
CC   human transcript.  This results in premature termination of the
CC   DNA-PKcs at amino acid 3160 (see AAW56642).  Oligonucleotide probes
CC   (see AAV30194 and AAV30195) are provided that precisely span the SCID
CC   determinant region of the DNA-PKcs gene, and which are diagnostic
CC   for the normal and SCID alleles, respectively.  Methods are also
CC   provided for identifying for differentiating SCID homozygotes,
CC   heterozygotes and normal horses.
XX
SQ   Sequence 11878 BP; 3586 A; 2323 C; 2668 G; 3301 T; 0 other;


  Query Match             100.0%;  Score 18;  DB 19;  Length 11878;
  Best Local Similarity   100.0%;  Pred. No. 45;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||||||||||||||||
Db   6429 tgcagcagcggcagaagt 6446


RESULT   6
AAV30198
ID   AAV30198 standard; DNA; 11883 BP.
XX
AC   AAV30198;
XX
DT   14-SEP-1998  (first entry)
XX
DE   Protein kinase catalytic subunit gene.
XX
KW   Severe combined immunodeficiency disease; SCID; horse; diagnosis;
KW   DNA-dependent protein kinase; ds.
XX
OS   Equus caballus.
XX
PN   WO9821367-A1.
XX
PD   22-MAY-1998.
XX
```

```
PF   14-NOV-1997;   97WO-US21066.
XX
PR   15-NOV-1996;   96US-0031261.
XX
PA   (TEXA ) UNIV TEXAS SYSTEM.
XX
PI   Meeks K;
XX
DR   WPI; 1998-297967/26.
DR   P-PSDB; AAW56642.
XX
PT   DNA-dependent protein kinase catalytic subunit - useful for
PT   determining equine severe combined immunodeficiency alleles
XX
PS   Claim 1; Page 39-44; 98pp; English.
XX
CC   This isolated DNA molecule encodes a DNA-dependent protein kinase
CC   catalytic subunit (DNA-PKcs, see AAW56642) found in Arabian horses.
CC   A claimed method of identifying an Arabian horse that is a carrier
CC   of equine severe combined immunodeficiency (SCID) comprises
CC   determining whether the horse has a mutation in a SCID determinant
CC   region of the DNA-PKcs gene (see also AAV30196 and AAV30197).  Sequence
CC   analysis of DNA-PKsc genes from normal and SCID equine fibroblasts
CC   shows that a 5 bp deletion is present in SCID foals at a site that
CC   corresponds to nucleotide 9454 of the 12,381 nucleotide coding
CC   sequence of the human transcript.  This results in premature
CC   termination of the DNA-PKsc at amino acid 3160.  Oligonucleotide
CC   probes (see AAV30194 and AAV30195) are provided that precisely span the
CC   SCID determinant region of the DNA-PKcs gene, and which are
CC   diagnostic for the normal and SCID alleles, respectively.  Methods
CC   are also provided for identifying for differentiating SCID
CC   homozygotes, heterozygotes and normal horses.  A claimed plasmid
CC   contains the DNA-PKcs DNA and regulatory elements necessary for
CC   expression of the DNA in a recombinant cell.
XX
SQ   Sequence 11883 BP; 3588 A; 2324 C; 2669 G; 3302 T; 0 other;


  Query Match             100.0%;  Score 18;  DB 19;  Length 11883;
  Best Local Similarity   100.0%;  Pred. No. 45;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||||||||||||||||
Db   6432 tgcagcagcggcagaagt 6449


RESULT   7
AAN70094
ID   AAN70094 standard; DNA; 876 BP.
XX
AC   AAN70094;
XX
DT   24-APR-1991  (first entry)
XX
DE   Sequence of region in front of the neutral protease gene in the
DE   Bacillus subtilis strain BGSC 1A1.
XX
KW   Promoter; Bacillus subtilis; pSM127; ss.
XX
OS   Bacillus subtilis.
XX
FH   Key             Location/Qualifiers
FT   misc_diff       452..517
FT                   /*tag= a
FT                   note= "absent from nprR2"
FT   repeat_region   510..577
FT                   /*tag=  b
FT                   /rpt_type= INVERTED
FT                   /note= "3 pairs of inverted repeats"
FT   promoter        571..599
FT                   /*tag=  c
FT                   /label= P1
FT   promoter        574..598
FT                   /*tag=  d
FT                   /label= P2
XX
PN   EP213085-A.
XX
PD   04-MAR-1987.
XX
PF   30-JUN-1986;   86EP-0830185.
XX
PR   10-JUL-1985;   85IT-0021507.
XX
PA   (ENIE ) ENIRICERCHE SPA.
XX
PI   Toma S,  Del Bue M,  Grandi G,  Mele A;
XX
DR   WPI; 1987-058527/09.
XX
PT   Recombinant DNA cloning vector - used for the expression and
PT   secretion of heterologous gene prods. in Bacillus subtilis
XX
PS   Disclosure; Fig 2B; 17pp; English.
XX
CC   The nprR2 region is differentiated from that of B.subtilis BGSC 1A1
CC   in that there is a deletion of 66 base pairs at a distance of 124
CC   bp from the origin of the gene, (see AAN70093,N70094).A new cloning
CC   vector based on pSM127 is claimed in which the heterologous DNA is under
CC   the control of the promoter regions, the centre of recognition of
CC   the ribosomes, nprR2 and the signal sequence of the neutral protease.
XX
SQ   Sequence 876 BP; 289 A; 185 C; 159 G; 243 T; 0 other;


  Query Match             91.1%;  Score 16.4;  DB 8;  Length 876;
  Best Local Similarity   94.4%;  Pred. No. 1.9e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||||||||||||| |||
Db     63 tgcagcagcggcagcagt 80


RESULT   8
AAV22141
ID   AAV22141 standard; cDNA; 36519 BP.
XX
AC   AAV22141;
XX
DT   17-AUG-1998  (first entry)
XX
DE   Chimpanzee adenovirus C68 genomic sequence.
XX
KW   genomic sequence; viral genome; virus; gene therapy; treatment;
KW   acquired defects; inherited defects; genetic engineering; vector;
KW   in vitro production; recombinant protein; ds.
XX
OS   Chimpanzee adenovirus C68.
XX
FH   Key             Location/Qualifiers
FT   CDS             480..1521
FT                   /*tag=  a
FT                   /note= "E1A gene"
FT   CDS             1560..3956
FT                   /*tag=  b
FT                   /note= "E1B gene"
FT   CDS             complement(23370..21787)
FT                   /*tag=  c
FT                   /note= "E2A gene"
FT   CDS             complement(10346..3957)
FT                   /*tag=  d
FT                   /note= "E2B gene"
FT   CDS             26806..31877
FT                   /*tag=  e
```

```
FT                   /note= "E3 gene"
FT   CDS             complement(36193..33486)
FT                   /*tag=  f
FT                   /note= "E4 gene"
FT   CDS             10823..13817
FT                   /*tag=  g
FT                   /note= "L1 gene"
FT   CDS             13884..17431
FT                   /*tag=  h
FT                   /note= "L2 gene"
FT   CDS             17480..21804
FT                   /*tag=  i
FT                   /note= "L3 gene"
FT   CDS             23399..27439
FT                   /*tag=  j
FT                   /note= "L4 gene"
FT   CDS             32134..33502
FT                   /*tag=  k
FT                   /note= "L5 gene"
XX
PN   WO9810087-A1.
XX
PD   12-MAR-1998.
XX
PF   04-SEP-1997;   97WO-US15694.
XX
PR   06-SEP-1996;   96US-0024700.
XX
PA   (UYPE-) UNIV PENNSYLVANIA.
XX
PI   Farina SF,  Fisher KJ,  Wilson JM;
XX
DR   WPI; 1998-193635/17.
XX
PT   Chimpanzee adenovirus vector - useful in gene therapy and genetic
PT   engineering in general
XX
PS   Disclosure; Pages 70-89; 116pp; English.
XX
CC   The sequence is that of the chimpanzee adenovirus C68. It may be
CC   used in the production of a vector comprising it and a selected
CC   heterologous gene operatively linked to regulatory sequences
CC   directing its expression in a heterologous cell.
CC   Such recombinant adenoviruses are useful as vectors in gene
CC   therapy and genetic engineering in general, especially for
CC   the treatment of aquired or genetically inherited defects.
CC   The viruses are also useful for the in vitro production of
CC   recombinant proteins of interest.
XX
SQ   Sequence 36519 BP; 7809 A; 10790 C; 10623 G; 7297 T; 0 other;


  Query Match             91.1%;  Score 16.4;  DB 19;  Length 36519;
  Best Local Similarity   94.4%;  Pred. No. 2.2e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          |||||||||| |||||||
Db  31662 tgcagcagcgccagaagt 31679


RESULT   9
AAA14833/c
ID   AAA14833 standard; DNA; 20 BP.
XX
AC   AAA14833;
XX
DT   08-AUG-2000  (first entry)
XX
DE   PCR primer HG52.165F for human G-protein coupled receptor HG52 cDNA.
XX
KW   Human; G-protein coupled receptor; HG52; immune system; thrombin;
KW   fibrinogen; fibrin; clotting factor; procoagulant; platelet activation;
KW   chemotaxis; mitogenesis; PCR primer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200020438-A1.
XX
PD   13-APR-2000.
XX
PF   29-SEP-1999;   99WO-US22634.
XX
PR   02-OCT-1998;   98US-0102958.
XX
PA   (MERI ) MERCK & CO INC.
XX
PI   Liu Q,  McDonald TP,  Wang R;
XX
DR   WPI; 2000-317696/27.
XX
PT   New recombinant DNA encoding a G-protein coupled receptor designated
PT   HG52 is useful to find modulators of thrombin effects and shows
PT   homology to thrombin receptors  -
XX
PS   Example 1; Page 23; 39pp; English.
XX
CC   PCR primers AAA14831-44 were used to amplify cDNA encoding a human
CC   G-protein coupled receptor designated HG52. HG52 is a member of
CC   the rhodopsin family. The HG52 RNA is widely expressed in humans as
CC   a transcript of about 4.5. kb, especially in cells of the immune
CC   system. The HG52 DNA can be used in chromosomal mapping studies, and
CC   to identify individuals carrying a disease-carrying gene. Agonists
CC   and antagonists of HG52 will be useful as modulators of the effects
CC   of thrombin, including conversion of fibrinogen to fibrin in plasma,
CC   activation of clotting factors V, VIII, XIII and protein C, the
CC   procoagulant function of platelets and endothelial cells, stimulation
CC   of platelet activation, chemotaxis of monocytes and lymphocytes, and
CC   mitogenesis of lymphocytes and mesenchymal cells such as vascular
CC   smooth muscle cells, fibroblasts and epithelial cells.
XX
SQ   Sequence 20 BP; 1 A; 8 C; 4 G; 7 T; 0 other;


  Query Match             85.6%;  Score 15.4;  DB 21;  Length 20;
  Best Local Similarity   94.1%;  Pred. No. 4.3e+02;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 gcagcagcggcagaagt 18
          | ||||||||||||||||
Db     20 GAAGCAGCGGCAGAAGT 4


RESULT  10
AAX06966
ID   AAX06966 standard; DNA; 39 BP.
XX
AC   AAX06966;
XX
DT   10-MAY-1999  (first entry)
XX
DE   Bacillus agaradhaerens primer xyloglucanase.upper.SacII.
XX
KW   Xyloglucanase; detergent; PCR; primer; ss.
XX
OS   Synthetic.
OS   Bacillus agaradhaerens.
XX
PN   WO9902663-A1.
XX
PD   21-JAN-1999.
XX
PF   01-JUL-1998;   98WO-DK00290.
XX
PR   24-OCT-1997;   97DK-0001213.
PR   07-JUL-1997;   97DK-0000822.
```

```
XX
PA   (NOVO ) NOVO-NORDISK AS.
XX
PI   Bjornvad ME,  Jorgensen PL,  Outtrup H,  Schuelein M;
XX
DR   WPI; 1999-120866/10.
XX
PT   New enzyme preparation comprising a xyloglucanase with an activity
PT   of 50 % at pH 7 - useful for improving the properties of cellulosic
PT   fibres, yarn, (non)woven fabric, and ratting hemp, jute, flax and
PT   linen fibres
XX
PS   Example 6; Page 63; 87pp; English.
XX
CC   Primers xyloglucanase.upper.SalII and xyloglucosidase.lower.NotI
CC   (see AAX06967) were used for the PCR amplification of Bacillus
CC   agaradhaerens NCIMB 40482 chromosomal DNA.  The primers incorporate
CC   SalII and NotI sites into the amplified DNA.  The PCR product was
CC   ligated into plasmid pMOL944, a pUB110 derivative containing
CC   elements making it propagatable in Bacillus subtilis.  The B.
CC   agarahaerens xyloglucanase (see AAW88463) was expressed in
CC   transformed B. subtilis PL2954 cells.  The enzyme is useful for
CC   improving the properties of cellulosic fabrics and in detergent
CC   compositions.
XX
SQ   Sequence 39 BP; 9 A; 12 C; 10 G; 8 T; 0 other;


  Query Match             85.6%;  Score 15.4;  DB 20;  Length 39;
  Best Local Similarity   94.1%;  Pred. No. 4.4e+02;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 tgcagcagcggcagaag 17
           ||||||  ||||||||||
Db       6 tgcagccgcggcagaag 22


RESULT  11
AAA35346
ID   AAA35346 standard; DNA; 240 BP.
XX
AC   AAA35346;
XX
DT   25-JUL-2000  (first entry)
XX
DE   Myrtaceae microsatellite sequence scu073TT including flanking regions.
XX
KW   Myrtaceae; microsatellite; isolation; genotyping; plant; tea tree;
KW   breeding; Melaleuca alternifolia; broad-spectrum germicidal oil;
KW   pharmaceutical; cosmetic; identification; detection; ds.
XX
OS   Myrtaceae sp.
XX
PN   WO200017341-A1.
XX
PD   30-MAR-2000.
XX
PF   23-SEP-1999;   99WO-AU00820.
XX
PR   23-SEP-1998;   98AU-0006099.
PR   16-FEB-1999;   99AU-0008718.
XX
PA   (BUSI-) BUSINESS & RES MANAGEMENT PTY LTD.
XX
PI   Rossetto M,  McLauchlan A,  Harriss FCL,  Henry RJ,  Baverstock PR;
PI   Lee LS,  Maguire TL,  Edwards KJ;
XX
DR   WPI; 2000-292840/25.
XX
PT   Isolating microsatellites from Myrtaceae, useful for genotyping,
PT   particularly in breeding programs for tea tree, by reacting plant
PT   nucleic acid with immobilized oligonucleotides -
XX
PS   Claim 8; Fig 35; 100pp; English.
XX
CC   A method has been developed of isolating a microsatellite (MS) from
CC   nucleic acid extract of a plant of Myrtaceae family. The method
CC   comprises: (i) treating the extract with one or more immobilised,
CC   single-stranded oligonucleotides (ON) having a consensus MS repeat
CC   sequence (MSRS) or its complement; (ii) washing under specified
CC   stringency conditions; (iii) eluting nucleic acid bound to ON; and
CC   (iv) sequencing the eluted nucleic acids to identify those containing
CC   an MSRS. Microsatellites (MS) isolated by the method, specifically
CC   from Melaleuca alternifolia (the tea tree, a source of a broad-spectrum
CC   germicidal oil, useful in pharmaceuticals and cosmetics), are useful as
CC   genotyping markers, particularly for breeding plants that produce the
CC   oil in higher yield or of better quality. Primers based on MS are
CC   useful for both inter- and intra-species genotyping. The selected
CC   washing conditions improve efficiency of recovery of microsatellites
CC   (MS) and reduce the number of washing stages required. Particularly
CC   about 86% of recovered sequence contain an MS repeat sequence,
CC   compared with 50-70% when the conventional washing procedure is
CC   followed. AAA35313 to AAA35357, and AAA35562 to AAA35575 represent
CC   nucleotide sequences from the present invention which contain
CC   microsatellite sequences. AAA35358 to AAA35561 represent oligonucleotide
CC   PCR primers used for identifying Myrtaceae microsatellite sequences.
XX
SQ   Sequence 240 BP; 52 A; 63 C; 67 G; 54 T; 4 other;


  Query Match             85.6%;  Score 15.4;  DB 21;  Length 240;
  Best Local Similarity   94.1%;  Pred. No. 4.8e+02;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 tgcagcagcggcagaag 17
           ||||||||||||||  ||
Db     107 tgcagcagcggcagcag 123


RESULT  12
AAT18914
ID   AAT18914 standard; DNA; 303 BP.
XX
AC   AAT18914;
XX
DT   17-JAN-1997  (first entry)
XX
DE   DNA encoding spider dragline variant, DP-1B.16 monomer.
XX
KW   Spider; dragline protein; variant;  monomer; polymer;
KW   fibre forming region; Spidroin 1; Nephila clavipes; DP1; mimic;
KW   DP-1A analogue; fibre; high tensile strength; elasticity; clothing;
KW   rope; surgical suture; implant; reinforcement; film; coating; ss.
XX
OS   Synthetic.
XX
PN   WO9429450-A2.
XX
PD   22-DEC-1994.
XX
PF   15-JUN-1994;   94WO-US06689.
XX
PR   15-JUN-1993;   93US-0077600.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Fahnestock SR;
XX
DR   WPI; 1995-036479/05.
DR   P-PSDB; AAR99056.
XX
PT   New synthetic variants of spider dragline protein - for making
PT   fibres useful as clothing, surgical silk, plastic reinforcement
PT   etc., also related DNA, vectors and transformed cells
XX
PS   Claim 8; Page 122; 168pp; English.
```

```
XX
CC   This sequence encodes a synthetic spider dragline variant monomer,
CC   DP-1B.16.  The sequence of the DP-1B.16 polymer is given in AAR99057.
CC   The polypeptide monomer is a variant based on a consensus sequence
CC   derived from the fibre forming regions of spider dragline protein,
CC   esp. the natural protein 1 (Spidroin 1) from Nephila clavipes.  This
CC   DNA sequence may be used in the recombinant production of the
CC   variant protein in a recombinant host, e.g. E. coli or Bacillus
CC   subtilis.  Synthetic analogues of DP1 were designed to mimic the
CC   repeating consensus sequence of the natural protein and the pattern
CC   of variation among individual repeats.  DP-1B analogues were designed
CC   by reordering the four repeats within the monomer of DP1A.  This
CC   monomer exhibits all of the regularities of (1)-(5) below.  In addition,
CC   it exhibits a regularity of the natural sequence which is not shared by
CC   DP-1A, namely that a repeat in which both GYG and GRG are deleted is
CC   generally preceded by a repeat lacking the entire poly-alanine repeat,
CC   with one intervening repeat.  The sequence of DP-1B matches the natural
CC   sequence more closely over a more extended segment than does DP-1A.  The
CC   individual repeats differ from the consensus according to the pattern:
CC          (1) the poly-alanine sequence varies in length from 0-7
CC          residues; (2) when the entire poly-alanine sequence is deleted,
CC   so also is the surrounding sequence encompassing AGRGGLGGQGAGAnGG;
CC          (3) aside from the poly-alanine sequence, deletions usually
CC   encompass integral multiples of three consecutive residues;
CC          (4) deletion of GYG is generally accompanied by deletion of GRG
CC   in the same sequence; and
CC          (5) a repeat in which the entire poly-alanine sequence is
CC   deleted is generally preceded by a repeat containing six alanine
CC   residues.
CC   The proteins may be used to produce fibres of high tensile strength and
CC   elasticity, suitable for clothing, rope, surgical sutures, biomaterials
CC   for implants; plastic reinforcements, films, coatings, etc.
XX
SQ   Sequence 303 BP; 34 A; 67 C; 140 G; 62 T; 0 other;


  Query Match             85.6%;  Score 15.4;  DB 16;  Length 303;
  Best Local Similarity   94.1%;  Pred. No. 4.8e+02;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaag 17
          ||||||||||||||| ||
Db    249 tgcagcagcggcagcag 265


RESULT  13
AAC12718/c
ID   AAC12718 standard; cDNA; 328 BP.
XX
AC   AAC12718;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 16793.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 16793; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 328 BP; 60 A; 112 C; 97 G; 56 T; 3 other;


  Query Match             85.6%;  Score 15.4;  DB 21;  Length 328;
  Best Local Similarity   94.1%;  Pred. No. 4.8e+02;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 gcagcagcggcagaagt 18
          | ||||||||||||||||
Db    258 GAAGCAGCGGCAGAAGT 242


RESULT  14
AAC03564
ID   AAC03564 standard; cDNA; 341 BP.
XX
AC   AAC03564;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 3562.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   P-PSDB; AAG03558.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 3562; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. An ORF has been identified within the
CC   sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
```

```
CC   derived from 30 different tissues. EST sequences usually correspond
CC   mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC   often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC   well suited for isolating cDNA sequences derived from the 5' ends of
CC   mRNAs and even in those cases where longer cDNA sequences have been
CC   obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC   mRNAs with intact 5' ends and can therefore be used to obtain full length
CC   cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC   gene therapy and chromosome mapping procedures. They are used to obtain
CC   upstream regulatory sequences and to design expression and secretion
CC   vectors.
XX
SQ   Sequence 341 BP; 68 A; 90 C; 115 G; 66 T; 2 other;


  Query Match             85.6%;  Score 15.4;  DB 21;  Length 341;
  Best Local Similarity   94.1%;  Pred. No. 4.8e+02;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 gcagcagcggcagaagt 18
          |||||||||||||| |||
Db    216 gcagcagcggcagcagt 232


RESULT  15
AAC79295
ID   AAC79295 standard; cDNA; 422 BP.
XX
AC   AAC79295;
XX
DT   05-FEB-2001  (first entry)
XX
DE   Human lung tumour-specific cDNA #248.
XX
KW   Lung tumour protein; lung cancer; cytostatic; vaccine; ss.
XX
OS   Homo sapiens.
XX
PN   WO200060077-A2.
XX
PD   12-OCT-2000.
XX
PF   30-MAR-2000; 2000WO-US08560.
XX
PR   02-APR-1999;   99US-0285323.
PR   09-AUG-1999;   99US-0370838.
PR   30-DEC-1999;   99US-0476235.
PR   03-MAR-2000; 2000US-0518809.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Reed SG,  Lodes MJ,  Mohamath R,  Secrist H;
XX
DR   WPI; 2000-638466/61.
XX
PT   Novel lung tumor polypeptides and polynucleotides, useful for
PT   detecting, monitoring or treating cancer, especially lung cancer -
XX
PS   Claim 3; Page 227; 243pp; English.
XX
CC   The present sequence is given in a specification relating to compounds
CC   for therapy and diagnosis of lung cancer. Polypeptides comprising at
CC   least an immunogenic part of a lung tumour protein are disclosed.
CC   The polypeptides are useful for inhibiting the development of cancer,
CC   especially lung cancer. Samples of T cells expressing the polypeptides
CC   may be used to inhibit the development of cancer. The polypeptides are
CC   also useful for detecting and monitoring the progression of cancer,
CC   especially lung cancer.
XX
SQ   Sequence 422 BP; 103 A; 113 C; 163 G; 42 T; 1 other;


  Query Match             85.6%;  Score 15.4;  DB 21;  Length 422;
  Best Local Similarity   94.1%;  Pred. No. 4.9e+02;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 gcagcagcggcagaagt 18
          |||||||||||||| |||
Db    166 gcagcagcggcaggagt 182


Search completed: September 21, 2001, 02:42:38
Job time: 11454 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: September 20, 2001, 23:27:52 ; Search time 4309.39 Seconds
(without alignments)
39.484 Million cell updates/sec

Title: US-09-138-735-8
Perfect score: 18
Sequence: 1 tgcagcagcggcagaagt 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est75:*
137: gb_est76:*
138: gb_est77:*
139: gb_est78:*
140: gb_est79:*
141: gb_est80:*
142: gb_est81:*
143: gb_est82:*
144: gb_est83:*
145: gb_est84:*
146: gb_est85:*
147: gb_est86:*
148: gb_est87:*
149: gb_est88:*
150: gb_est89:*
151: gb_est90:*
152: gb_est99:*
153: gb_est100:*
154: gb_est101:*
155: gb_est102:*
156: gb_est103:*
157: gb_est104:*
158: gb_est105:*
159: gb_est106:*
160: gb_est67:*
161: gb_est68:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_esthum29:*
177: em_esthum30:*
178: em_esthum31:*
179: em_esthum32:*
180: em_esthum33:*
181: em_estom3:*
182: em_estpl11:*
183: em_estro21:*
184: em_estro22:*
185: em_estro23:*
186: em_htc:*
187: gb_est107:*
188: gb_est108:*
189: gb_est109:*
190: gb_est110:*
191: gb_est111:*
192: gb_htc:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_other:*
207: em_gss_pln1:*
208: em_gss_pln2:*
209: em_gss_pro:*
210: em_gss_rod1:*
211: em_gss_rod2:*
212: em_gss_rod3:*
213: em_gss_rod4:*
214: em_gss_rod5:*
215: em_gss_vrt1:*
216: em_gss_vrt2:*
217: em_gss_vrt3:*
218: gb_gss1:*
219: gb_gss2:*
220: gb_gss3:*
221: gb_gss4:*
222: gb_gss5:*
223: gb_gss6:*
224: gb_gss7:*
225: gb_gss8:*
226: gb_gss9:*
227: gb_gss10:*
228: gb_gss11:*
229: gb_gss12:*
230: gb_gss13:*
231: gb_gss14:*
232: gb_gss15:*
233: gb_gss16:*
234: gb_gss17:*
235: gb_gss18:*
236: gb_gss19:*
237: gb_gss20:*
238: gb_gss21:*
239: gb_gss22:*
240: gb_gss23:*
241: gb_gss24:*
242: gb_gss25:*
243: gb_gss26:*
244: gb_gss27:*
245: gb_gss28:*
246: gb_gss29:*
247: gb_gss30:*
248: gb_gss31:*
249: gb_gss32:*
250: gb_gss33:*
251: gb_gss34:*
252: em_gss_inv4:*
253: em_gss_rod6:*
254: em_gss_rod7:*
255: em_gss_rod8:*
256: gb_gss35:*
257: gb_gss36:*
258: gb_gss37:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	17	94.4	524	138	BE665288	BE665288 153467 MA
2	17	94.4	563	138	BE665781	BE665781 154938 MA
3	17	94.4	658	238	AZ086647	AZ086647 RPCI-23-2
4	17	94.4	725	248	AZ705723	AZ705723 RPCI-23-2
c 5	16.4	91.1	201	192	AK019382	AK019382 Mus muscu
c 6	16.4	91.1	203	25	AV158967	AV158967 AV158967
c 7	16.4	91.1	208	161	BB591907	BB591907 BB591907
c 8	16.4	91.1	220	161	BB595622	BB595622 BB595622
c 9	16.4	91.1	240	161	BB563512	BB563512 BB563512
c 10	16.4	91.1	307	165	BE244644	BE244644 TCBAP2E00
c 11	16.4	91.1	337	23	AI642272	AI642272 vn76a01.y
c 12	16.4	91.1	358	105	AL362372	AL362372 AL362372
c 13	16.4	91.1	418	138	BE651324	BE651324 UI-M-BH3-
14	16.4	91.1	421	232	AQ680611	AQ680611 HS_5496_A
c 15	16.4	91.1	424	144	BF116881	BF116881 uy92d12.y
c 16	16.4	91.1	425	15	AI035972	AI035972 ub50c11.r
c 17	16.4	91.1	442	13	AA867357	AA867357 vx97c08.r
c 18	16.4	91.1	443	154	BG516264	BG516264 EtESTed56
c 19	16.4	91.1	445	230	AQ589778	AQ589778 HS_2117_B
c 20	16.4	91.1	459	154	BG466312	BG466312 EtESTed51
c 21	16.4	91.1	471	10	AA638912	AA638912 vn76a01.r
c 22	16.4	91.1	475	117	AW530300	AW530300 UI-R-BU0-
c 23	16.4	91.1	484	114	AW322968	AW322968 uo54f09.y
24	16.4	91.1	489	239	AZ183201	AZ183201 SP_0191_A
c 25	16.4	91.1	502	113	AW230793	AW230793 uo67h04.y
c 26	16.4	91.1	505	114	AW320029	AW320029 un13e11.x
c 27	16.4	91.1	528	222	FR0034786	AL122318 Fugu rubr
c 28	16.4	91.1	531	139	BE744774	BE744774 601573161
c 29	16.4	91.1	552	165	BE290024	BE290024 601089630
30	16.4	91.1	554	240	AZ220432	AZ220432 Sheared D
31	16.4	91.1	558	239	AZ154969	AZ154969 SP_0040_A
c 32	16.4	91.1	569	231	AQ647766	AQ647766 RPCI93-Dp
c 33	16.4	91.1	577	166	BE304046	BE304046 601085519
34	16.4	91.1	577	251	AZ866126	AZ866126 2M0176F01
35	16.4	91.1	588	107	AU018956	AU018956 AU018956
36	16.4	91.1	593	222	FR0034779	AL122311 Fugu rubr
37	16.4	91.1	596	250	AZ859556	AZ859556 2M0165O12
38	16.4	91.1	615	122	AW918438	AW918438 EST349742
39	16.4	91.1	623	258	TA106E01P	AL459117 T. brucei
c 40	16.4	91.1	628	108	AU132503	AU132503 AU132503
41	16.4	91.1	636	149	BF491885	BF491885 AT28649.5
c 42	16.4	91.1	654	192	AK003238	AK003238 Mus muscu
43	16.4	91.1	655	149	BF486916	BF486916 AT21486.5
44	16.4	91.1	665	236	AQ946302	AQ946302 Sheared D
45	16.4	91.1	679	149	BF486412	BF486412 AT20711.5

ALIGNMENTS

```
RESULT   1
BE665288
LOCUS       BE665288      524 bp    mRNA            EST       08-SEP-2000
DEFINITION  153467 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BE665288
VERSION     BE665288.1  GI:10025706
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 524)
  AUTHORS   Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
            Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid
            ,W.W. and Keele,J.W.
  TITLE     Design and use of four pooled tissue normalized cDNA libraries for
            EST discovery in cattle
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 69  row: D  column: 6
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .524
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 4BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from day 20 and day 40
                     embryos."
BASE COUNT      106 a    145 c    153 g    120 t
ORIGIN


  Query Match             94.4%;  Score 17;  DB 138;  Length 524;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+03;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tgcagcagcggcagaag 17
           |||||||||||||||||
Db     286 TGCAGCAGCGGCAGAAG 302


RESULT   2
BE665781
LOCUS       BE665781      563 bp    mRNA            EST       08-SEP-2000
DEFINITION  154938 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BE665781
VERSION     BE665781.1  GI:10026372
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 563)
  AUTHORS   Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
            Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid
            ,W.W. and Keele,J.W.
  TITLE     Design and use of four pooled tissue normalized cDNA libraries for
            EST discovery in cattle
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 69  row: D  column: 7
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
```

```
     source          1. .563
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 4BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from day 20 and day 40
                     embryos."
BASE COUNT      116 a    158 c    157 g    131 t      1 others
ORIGIN


  Query Match             94.4%;  Score 17;  DB 138;  Length 563;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+03;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaag 17
          |||||||||||||||||
Db    286 TGCAGCAGCGGCAGAAG 302


RESULT   3
AZ086647
LOCUS       AZ086647      658 bp    DNA             GSS       08-MAY-2000
DEFINITION  RPCI-23-26F10.TJ RPCI-23 Mus musculus genomic clone RPCI-23-26F10,
            DNA sequence.
ACCESSION   AZ086647
VERSION     AZ086647.1  GI:7728381
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 658)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-23-26F10.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 26  row: F  column: 10
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .658
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-26F10"
                     /clone_lib="RPCI-23"
                     /sex="Female"
                     /lab_host="DH10B"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).   "
BASE COUNT      181 a    168 c    150 g    159 t
ORIGIN


  Query Match             94.4%;  Score 17;  DB 238;  Length 658;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+03;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaag 17
          |||||||||||||||||
Db    532 TGCAGCAGCGGCAGAAG 548


RESULT   4
AZ705723
LOCUS       AZ705723      725 bp    DNA             GSS       24-JAN-2001
DEFINITION  RPCI-23-234M21.TV RPCI-23 Mus musculus genomic clone RPCI-23-234M21
            , DNA sequence.
ACCESSION   AZ705723
VERSION     AZ705723.1  GI:12432498
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 725)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-23-234M21.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 234  row: M  column: 21
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .725
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-234M21"
                     /clone_lib="RPCI-23"
                     /sex="Female"
                     /lab_host="DH10B"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).   "
BASE COUNT      208 a    174 c    167 g    176 t
ORIGIN


  Query Match             94.4%;  Score 17;  DB 248;  Length 725;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+03;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaag 17
          |||||||||||||||||
```

```
Db    548 TGCAGCAGCGGCAGAAG 564


RESULT   5
AK019382/c
LOCUS       AK019382      201 bp    mRNA            HTC       08-FEB-2001
DEFINITION  Mus musculus 12 days embryo head cDNA, RIKEN full-length enriched
            library, clone:3010020L09, full insert sequence.
ACCESSION   AK019382
VERSION     AK019382.1  GI:12859557
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) 12 days embryo head cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:3010020L09.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods Enzymol. 303, 19-44 (1999)
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Kikuchi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T.,
            Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T.,
            Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Okazaki,Y., Muramatsu,M., Inoue,Y. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 201)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 10.0 and subtraction to
            Rot = 50.0. Second strand cDNA was prepared with the primer adapter
            of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES             Location/Qualifiers
     source          1. .201
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /db_xref="MGD:MGI:1899127"
                     /db_xref="MGD:MGI:1914915"
                     /clone="3010020L09"
                     /tissue_type="head"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="12 days embryo"
     CDS             13. .>201
                     /note="putative"
                     /codon_start=1
                     /protein_id="BAB31691.1"
                     /db_xref="GI:12859558"
                     /translation="MASLLQSERVLYLVQGEKKVRAPLSQLYFCRYCSELRSLECVSH
                     EVDSHYCPSCLENMPSAEA"
BASE COUNT       41 a     59 c     54 g     47 t
ORIGIN


  Query Match             91.1%;  Score 16.4;  DB 192;  Length 201;
  Best Local Similarity   94.4%;  Pred. No. 1.9e+03;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||| ||||||||||||
Db    112 TGCAGTAGCGGCAGAAGT 95


RESULT   6
AV158967/c
LOCUS       AV158967      203 bp    mRNA            EST       07-JUL-1999
DEFINITION  AV158967 Mus musculus head C57BL/6J 12-day embryo Mus musculus cDNA
            clone 3010020L09, mRNA sequence.
ACCESSION   AV158967
VERSION     AV158967.1  GI:5402602
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 203)
  AUTHORS   Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K.,
            Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T., Hara
            ,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M., Kawai,J.,
            Kikuchi,N., Kojima,Y., Matsuyama,T., Niitsuma,H., Oda,H., Owa,C.,
            Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara
            ,Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y., Tominaga,N.,
            Watanabe,S., Yagame,M., Yamamura,T., Yokota,T., Yoshino,M.,
            Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Chie Owa
            Genome Science Laboratory
            RIKEN
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-9145
            Fax: 81-298-36-9098
            Email: genome-res@rtc.riken.go.jp
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length cDNA
            (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
            Transcriptional sequencing: A method for DNA sequencing using RNA
```

polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.

```
FEATURES             Location/Qualifiers
     source          1. .203
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="3010020L09"
                     /clone_lib="Mus musculus head C57BL/6J 12-day embryo"
                     /sex="mixed"
                     /tissue_type="head"
                     /dev_stage="12-day embryo"
BASE COUNT       40 a     57 c     48 g     58 t
ORIGIN


  Query Match             91.1%;  Score 16.4;  DB 25;  Length 203;
  Best Local Similarity   94.4%;  Pred. No. 1.9e+03;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 tgcagcagcggcagaagt 18
           ||||| ||||||||||||
Db     114 TGCAGTAGCGGCAGAAGT 97


RESULT   7
BB591907/c
LOCUS       BB591907      208 bp    mRNA            EST       30-NOV-2000
DEFINITION  BB591907 RIKEN full-length enriched, 2 cells egg Mus musculus cDNA
            clone B020005D04 5', mRNA sequence.
ACCESSION   BB591907
VERSION     BB591907.1  GI:11488509
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 208)
  AUTHORS   Aizawa,K., Akahira,S., Akimura,T., Arai,A., Arakawa,T., Carninci,P.
            , Hanagaki,T., Hayatsu,N., Hiraoka,T., Hirozane,T., Hodoyama,Y.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kawai,J., Kojima,Y., Konno
            ,H., Kusakabe,M., Matsuyama,T., Miyazaki,A., Nakamura,M., Nishi,K.,
            Nomura,K., Numazaki,R., Okazaki,Y., Okido,T., Owa,C., Sakai,C.,
            Sakai,K., Sasaki,D., Sato,K., Shibata,K., Shibata,Y., Shinagawa,A.,
            Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Tanaka
            ,T., Toya,T., Watahiki,A., Yamamura,T., Yasunishi,A., Yoshida,K.,
            Yoshiki,A., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Aizawa,K. et al. 2000)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Genome Exploration Research Group, Life Science Tsukuba Center,
            Genome Science Laboratory
            The Institute of Physical and Chemical Research (RIKEN), Genomic
            Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: +81-298-36-9013
            Fax: +81-298-36-9098
            Email: genome-res@rtc.riken.go.jp,
            URL:http://genome.rtc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
```

Please visit our web site (http://genome.rtc.riken.go.jp) for further details.

```
FEATURES             Location/Qualifiers
     source          1. .208
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="B020005D04"
                     /clone_lib="RIKEN full-length enriched, 2 cells egg"
                     /tissue_type="egg"
                     /dev_stage="2 cells"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA
                     was cleaved with XhoI and BamHI. Vector: a modified
                     pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT       43 a     61 c     54 g     50 t
ORIGIN


  Query Match             91.1%;  Score 16.4;  DB 161;  Length 208;
  Best Local Similarity   94.4%;  Pred. No. 1.9e+03;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 tgcagcagcggcagaagt 18
           ||||| ||||||||||||
Db     132 TGCAGTAGCGGCAGAAGT 115


RESULT   8
BB595622/c
LOCUS       BB595622      220 bp    mRNA            EST       30-NOV-2000
DEFINITION  BB595622 RIKEN full-length enriched, 16 days embryo head Mus
            musculus cDNA clone C130002M02 5', mRNA sequence.
ACCESSION   BB595622
VERSION     BB595622.1  GI:11492224
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 220)
  AUTHORS   Aizawa,K., Akahira,S., Akimura,T., Arai,A., Arakawa,T., Carninci,P.
            , Hanagaki,T., Hayatsu,N., Hiraoka,T., Hirozane,T., Hodoyama,Y.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kawai,J., Kojima,Y., Konno
            ,H., Kusakabe,M., Matsuyama,T., Miyazaki,A., Nakamura,M., Nishi,K.,
            Nomura,K., Numazaki,R., Okazaki,Y., Okido,T., Owa,C., Sakai,C.,
            Sakai,K., Sasaki,D., Sato,K., Shibata,K., Shibata,Y., Shinagawa,A.,
            Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Tanaka
            ,T., Toya,T., Watahiki,A., Yamamura,T., Yasunishi,A., Yoshida,K.,
            Yoshiki,A., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Aizawa,K. et al. 2000)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Genome Exploration Research Group, Life Science Tsukuba Center,
            Genome Science Laboratory
            The Institute of Physical and Chemical Research (RIKEN), Genomic
            Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: +81-298-36-9013
            Fax: +81-298-36-9098
            Email: genome-res@rtc.riken.go.jp,
            URL:http://genome.rtc.riken.go.jp/
```

```
           Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
           ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
           Thermostabilization and thermoactivation of thermolabile enzymes by
           trehalose and its application for the synthesis of full length
           cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
            Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
           Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
           ,Y. and Hayashizaki,Y.
            Automated filtration-based high-throughput plasmid preparation
           system. Genome Res. 9 (5), 463-470 (1999)
            Carninci,P. and Hayashizaki,Y.
            High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
           19-44 (1999)
            Please visit our web site (http://genome.rtc.riken.go.jp) for
           further details.
FEATURES             Location/Qualifiers
     source          1. .220
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="C130002M02"
                     /clone_lib="RIKEN full-length enriched, 16 days embryo
                     head"
                     /sex="mixed"
                     /tissue_type="head"
                     /dev_stage="16 days embryo"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA
                     was cloned into the XhoI and BamHI sites. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I"
BASE COUNT       46 a     52 c     69 g     53 t
ORIGIN


  Query Match             91.1%;  Score 16.4;  DB 161;  Length 220;
  Best Local Similarity   94.4%;  Pred. No. 1.9e+03;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||| ||||||||||||
Db    127 TGCAGTAGCGGCAGAAGT 110


RESULT  9
BB563512/c
LOCUS       BB563512      240 bp    mRNA            EST       29-NOV-2000
DEFINITION  BB563512 RIKEN full-length enriched, 18 days embryo Mus musculus
            cDNA clone 1110001K06 5', mRNA sequence.
ACCESSION   BB563512
VERSION     BB563512.1  GI:11454404
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 240)
  AUTHORS   Aizawa,K., Akahira,S., Akimura,T., Arai,A., Arakawa,T., Carninci,P.
            , Hanagaki,T., Hayatsu,N., Hiraoka,T., Hirozane,T., Hodoyama,Y.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kawai,J., Kojima,Y., Konno
            ,H., Kusakabe,M., Matsuyama,T., Miyazaki,A., Nakamura,M., Nishi,K.,
            Nomura,K., Numazaki,R., Okazaki,Y., Okido,T., Owa,C., Sakai,C.,
            Sakai,K., Sasaki,D., Sato,K., Shibata,K., Shibata,Y., Shinagawa,A.,
            Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Tanaka
            ,T., Toya,T., Watahiki,A., Yamamura,T., Yasunishi,A., Yoshida,K.,
            Yoshiki,A., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Aizawa,K. et al. 2000)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Genome Exploration Research Group, Life Science Tsukuba Center,
            Genome Science Laboratory
            The Institute of Physical and Chemical Research (RIKEN), Genomic
            Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: +81-298-36-9013
            Fax: +81-298-36-9098
            Email: genome-res@rtc.riken.go.jp,
            URL:http://genome.rtc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
            Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
            Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
            Carninci,P. and Hayashizaki,Y.
            High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .240
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="1110001K06"
                     /clone_lib="RIKEN full-length enriched, 18 days embryo"
                     /dev_stage="18 days embryo"
                     /lab_host="SOLR"
                     /note="Site_1: XhoI; Site_2: SstI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCAATTAATTAATTAAACCCCCCCCCCC 3'].
                     cDNA was cleaved with XhoI and SstI."
BASE COUNT       56 a     54 c     66 g     64 t
ORIGIN


  Query Match             91.1%;  Score 16.4;  DB 161;  Length 240;
  Best Local Similarity   94.4%;  Pred. No. 1.9e+03;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||| ||||||||||||
Db    103 TGCAGTAGCGGCAGAAGT 86


RESULT  10
BE244644/c
LOCUS       BE244644      307 bp    mRNA            EST       15-NOV-2000
DEFINITION  TCBAP2E0065 Pediatric pre-B cell acute lymphoblastic leukemia
            Baylor-HGSC project=TCBA Homo sapiens cDNA clone TCBAP0065, mRNA
            sequence.
```

```
ACCESSION   BE244644
VERSION     BE244644.1  GI:9096386
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 307)
  AUTHORS   Wei,Y., Tsang,Y.T.M., Mei,G., Ku,J.M., Ali-Osman Jr.,F.R., Muzny,D.
            , Bouck,J., Gibbs,R.A. and Margolin,J.F.
  TITLE     Pediatric Leukemia cDNA Sequencing Project
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Dr. Judith F. Margolin
            Texas Children's Cancer Center and Human Genome Sequencing Center
            at Baylor College of Medicine
            1102 Bates, MC3-3320 Houston, TX 77030, USA
            Tel: 832-824-4536
            Fax: 832-825-4038
            Email: clones@txccc.org
            Seq primer: M13 primer.
FEATURES             Location/Qualifiers
     source          1. .307
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="TCBAP0065"
                     /clone_lib="Pediatric pre-B cell acute lymphoblastic
                     leukemia Baylor-HGSC project=TCBA"
                     /sex="male"
                     /tissue_type="leukopheresis"
                     /cell_type="pre-B cell"
                     /dev_stage="pediatric 2 years"
                     /lab_host="DH10B"
                     /note="Vector: lambda pSB; Site_1: BamHI; Site_2: EcoRI;
                     First strand cDNA was primed with an anchored
                     XhoI-oligo(dT) primer [5'GGAGGACTCGAGCGGCCGCAGGAGGAG(T)VN
                     3'; V=A,C,G; N=A,C,G,T] and then  dG tailed.  Second
                     strand was primed with a BamH1-dC primer
                     [5'AGAGAGCTCGGATCCGCGGCCGCAATAATAATAAT(C) 3'].
                     Double-stranded cDNA  was then digested with BamH1 and
                     XhoI and directionally cloned into  the BamH1 and SalI
                     sites of lambda pSB vector.  Library went through one
                     round of normalization.  Library was constructed by Wei
                     Yu"
BASE COUNT       38 a    112 c    107 g     49 t      1 others
ORIGIN
```

```
  Query Match             91.1%;  Score 16.4;  DB 165;  Length 307;
  Best Local Similarity   94.4%;  Pred. No. 2e+03;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          |||||| ||||||||||
Db    232 TGCAGCGGCGGCAGAAGT 215
```

```
RESULT  11
AI642272/c
LOCUS       AI642272      337 bp    mRNA            EST       29-APR-1999
DEFINITION  vn76a01.y1 Knowles Solter mouse blastocyst B1 Mus musculus cDNA
            clone IMAGE:1037832 5', mRNA sequence.
ACCESSION   AI642272
VERSION     AI642272.1  GI:4720747
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 337)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:577856
            This read is a RESEQUENCE of a previously sequenced mouse clone
            This read has been verified (found to hit its original self in the
            correct orientation)
            Seq primer: -40RP
            High quality sequence stop: 336.
FEATURES             Location/Qualifiers
     source          1. .337
                     /organism="Mus musculus"
                     /strain="B6D2 F1/J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1037832"
                     /clone_lib="Knowles Solter mouse blastocyst B1"
                     /tissue_type="blastocyst"
                     /dev_stage="embryo (pre-implantation)"
                     /lab_host="DH10B"
                     /note="Organ: embryo; Vector: pSPORT; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally from mRNA prepared
                     from 800 blastocysts.  Primer: SalI(dT):
                     5'-CGGTCGACCGTCGACCGTTTTTTTTTTTTTT-3'.  cDNAs  were
                     cloned into the NotI/SalI sites of a pSPORT vector (Life
                     Technologies).  Two different size selections:  B1 (larger
                     inserts) and B3."
BASE COUNT       54 a     92 c    117 g     72 t      2 others
ORIGIN
```

```
  Query Match             91.1%;  Score 16.4;  DB 23;  Length 337;
  Best Local Similarity   94.4%;  Pred. No. 2e+03;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          |||||| ||||||||||
Db    104 TGCAGTAGCGGCAGAAGT 87
```

```
RESULT  12
AL362372/c
LOCUS       AL362372      358 bp    mRNA            EST       04-AUG-2000
DEFINITION  AL362372 ICRFp 522 and 523 Mus musculus cDNA clone W9322B43 5',
            mRNA sequence.
ACCESSION   AL362372
VERSION     AL362372.1  GI:9694940
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 358)
  AUTHORS   Eickhoff,H., Schuchhardt,J., Ivanov,I., Meier-Ewert,S., O'Brien,J.,
            Malik,A., Tandon,N., Wolski,E., Rohlfs,E., Nyarsik,L., Reinhardt,R.
            , Nietfeld,W. and Lehrach,H.
  TITLE     Tissue gene expression analysis using arrayed normalized cDNA
            libraries
  JOURNAL   Genome Res. (2000) In press
COMMENT     Contact: MPIMG
            Abt.Lehrach
            Max Planck Institut Fuer Molekulare Genetik
            Ihnestrasse 73, Berlin, 14195 Germany
            The cDNA libraries ICRFp 522 and 523 were normalized with
            oligonucleotide fingerprinting, resulting in a unique subset of
            5376 cDNA clones.
```

```
FEATURES             Location/Qualifiers
     source          1. .358
                     /organism="Mus musculus"
                     /strain="Black 6"
                     /db_xref="taxon:10090"
                     /clone="W9322B43"
                     /clone_lib="ICRFp 522 and 523"
                     /tissue_type="embryo"
                     /dev_stage="9 and 12 pc embryo"
BASE COUNT       63 a     88 c    114 g     67 t     26 others
ORIGIN


  Query Match             91.1%;  Score 16.4;  DB 105;  Length 358;
  Best Local Similarity   94.4%;  Pred. No. 2e+03;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||| ||||||||||||
Db    120 TGCAGTAGCGGCAGAAGT 103


RESULT  13
BE651324/c
LOCUS       BE651324      418 bp    mRNA            EST       06-SEP-2000
DEFINITION  UI-M-BH3-asy-c-12-0-UI.r1 NIH_BMAP_M_S4 Mus musculus cDNA clone
            UI-M-BH3-asy-c-12-0-UI 5', mRNA sequence.
ACCESSION   BE651324
VERSION     BE651324.1  GI:9977148
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 418)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It
            should be noted that Bento Soares is generating a small number of
            additional specialized non-redundant arrays of BMAP cDNAs whose
            availability will be considered under appropriate and limited
            collaborative arrangements
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .418
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-BH3-asy-c-12-0-UI"
                     /clone_lib="NIH_BMAP_M_S4"
                     /dev_stage="27-32 days"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NIH_BMAP_M_S4 library is a subtracted library of a series,
                     ultimately  derived from a mixture of individually tagged
                     normalized libraries from  ten regions of the mouse brain
                     (cerebellum, brain stems, olfactory  bulbs, hypothalamus,
                     cortex, amygdala, basal ganglia, pineal gland,  striatum,
                     hipoccampus) after a series of subtractions to reduce the
                     representation of cDNAs from which ESTs had already been
                     generated. The following serially subtracted libraries
                     were generated in this process: NIH_BMAP_M_S4,
                     NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1,
                     NIH_BMAP_M_S2, NIH_BMAP_M_S1.  The subtracted library
                     (NIH_BMAP_M_S4) was constructed as follows:  PCRamplified
                     cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and
                     NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived
                     was used as  a driver in a hybridization with a pool of
                     the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1
                     libraries in the form of  single-stranded circles.  The
                     remaining single-stranded circles  (subtracted library)
                     was purified by hydroxyapatite column chromatography,
                     converted to double-stranded circles and electroporated
                     into DH10B  bacteria (LifeTechnologies) to generate the
                     NIH_BMAP_M_S4 library.  This procedure has been previously
                     described (Bonaldo, Lennon and Soares, Genome Research
                     6:791-806, 1996)"
BASE COUNT      110 a    112 c    100 g     96 t
ORIGIN


  Query Match             91.1%;  Score 16.4;  DB 138;  Length 418;
  Best Local Similarity   94.4%;  Pred. No. 2e+03;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||| ||||||||||||
Db     92 TGCAGTAGCGGCAGAAGT 75


RESULT  14
AQ680611
LOCUS       AQ680611      421 bp    DNA             GSS       25-JUN-1999
DEFINITION  HS_5496_A1_B01_SP6E RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=1072 Col=1 Row=C, DNA sequence.
ACCESSION   AQ680611
VERSION     AQ680611.1  GI:5229415
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 421)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 1072  row: C  column: 1
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 421.
FEATURES             Location/Qualifiers
     source          1. .421
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=1072 Col=1 Row=C"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
```

```
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT      133 a     76 c     69 g    141 t      2 others
ORIGIN


  Query Match             91.1%;  Score 16.4;  DB 232;  Length 421;
  Best Local Similarity   94.4%;  Pred. No. 2e+03;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          |||||||||||| ||||||
Db     90 TGCAGCAGCGGTAGAAGT 107


RESULT  15
BF116881/c
LOCUS       BF116881      424 bp    mRNA            EST       29-DEC-2000
DEFINITION  uy92d12.y1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3667031 5'
            similar to TR:Q9QXP8 Q9QXP8 P62 DYNACTIN SUBUNIT. [1] ;, mRNA
            sequence.
ACCESSION   BF116881
VERSION     BF116881.1  GI:10986357
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 424)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            image.llnl.gov/image/html/iresources.shtml

            MGI:1427799
            Seq primer: -40RP from Gibco
            High quality sequence stop: 350.
FEATURES             Location/Qualifiers
     source          1. .424
                     /organism="Mus musculus"
                     /strain="C57/B6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3667031"
                     /clone_lib="NCI_CGAP_Mam5"
                     /tissue_type="tumor, gross tissue"
                     /dev_stage="7 months"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigators
                     providing samples: Lothar Hennighausen/Robin Humphreys,
                     NIH"
BASE COUNT      111 a    115 c    102 g     96 t
ORIGIN


  Query Match             91.1%;  Score 16.4;  DB 144;  Length 424;
  Best Local Similarity   94.4%;  Pred. No. 2e+03;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          |||||| ||||||||||||
Db    107 TGCAGTAGCGGCAGAAGT 90


Search completed: September 20, 2001, 23:28:03
Job time: 25543 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: September 21, 2001, 02:16:29 ; Search time 164.23 Seconds
(without alignments)
20.749 Million cell updates/sec

Title: US-09-138-735-8
Perfect score: 18
Sequence: 1 tgcagcagcggcagaagt 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn1_7/ptodata/1/ina/5A_COMB.seq:*
2: /cgn1_7/ptodata/1/ina/5B_COMB.seq:*
3: /cgn1_7/ptodata/1/ina/6A_COMB.seq:*
4: /cgn1_7/ptodata/1/ina/6B_COMB.seq:*
5: /cgn1_7/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn1_7/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	18	100.0	18	1	US-08-480-917-8	Sequence 8, Appli
c	2	18	100.0	3402	1	US-08-480-917-1	Sequence 1, Appli
	3	18	100.0	11873	2	US-08-970-269A-32	Sequence 32, Appl
	4	18	100.0	11878	2	US-08-970-269A-31	Sequence 31, Appl
	5	18	100.0	11883	2	US-08-970-269A-28	Sequence 28, Appl
	6	16.4	91.1	36519	3	US-08-923-137-2	Sequence 2, Appli
	7	15.4	85.6	39	4	US-09-110-959A-13	Sequence 13, Appl
	8	15.4	85.6	1276	4	US-09-177-325-2	Sequence 2, Appli
	9	15.4	85.6	2000	1	US-07-923-739-1	Sequence 1, Appli
c	10	15.4	85.6	2214	3	US-08-864-038A-1	Sequence 1, Appli
c	11	15.4	85.6	3331	3	US-08-864-038A-2	Sequence 2, Appli
c	12	15.4	85.6	3331	3	US-08-864-038A-4	Sequence 4, Appli
	13	15.4	85.6	5197	1	US-08-131-365B-53	Sequence 53, Appl
	14	15.4	85.6	5197	2	US-08-668-123-53	Sequence 53, Appl
	15	15	83.3	1869	3	US-08-934-494-1	Sequence 1, Appli
	16	15	83.3	1869	3	US-09-143-068-1	Sequence 1, Appli
	17	14.8	82.2	38	4	US-08-470-369-24	Sequence 24, Appl
	18	14.8	82.2	38	6	5304637-24	Patent No. 5304637
c	19	14.8	82.2	1033	2	US-08-414-657D-19	Sequence 19, Appl
c	20	14.8	82.2	1851	2	US-08-414-657D-20	Sequence 20, Appl
c	21	14.8	82.2	2232	1	US-08-179-481-97	Sequence 97, Appl
c	22	14.8	82.2	2464	1	US-07-863-169A-4	Sequence 4, Appli
c	23	14.8	82.2	2464	2	US-08-429-964-4	Sequence 4, Appli
c	24	14.8	82.2	2464	3	US-07-935-087-4	Sequence 4, Appli
c	25	14.8	82.2	2464	5	PCT-US93-08062-4	Sequence 4, Appli
c	26	14.8	82.2	2603	1	US-08-179-481-1	Sequence 1, Appli
c	27	14.8	82.2	2716	1	US-08-647-484-1	Sequence 1, Appli
c	28	14.8	82.2	2716	1	US-08-647-484-3	Sequence 3, Appli
c	29	14.8	82.2	2716	1	US-08-647-481-1	Sequence 1, Appli
c	30	14.8	82.2	2716	1	US-08-647-481-3	Sequence 3, Appli
c	31	14.8	82.2	2716	1	US-08-430-033A-1	Sequence 1, Appli
c	32	14.8	82.2	2716	1	US-08-430-033A-3	Sequence 3, Appli
c	33	14.8	82.2	2716	5	PCT-US96-05792-1	Sequence 1, Appli
c	34	14.8	82.2	2716	5	PCT-US96-05792-3	Sequence 3, Appli
c	35	14.8	82.2	5183	3	US-09-039-555B-18	Sequence 18, Appl
	36	14.8	82.2	6942	2	US-08-460-309-3	Sequence 3, Appli
	37	14.8	82.2	6942	2	US-08-125-077-3	Sequence 3, Appli
	38	14.4	80.0	19	1	US-08-410-540-5	Sequence 5, Appli
c	39	14.4	80.0	110	1	US-07-668-648-11	Sequence 11, Appl
c	40	14.4	80.0	110	2	US-08-429-998-11	Sequence 11, Appl
c	41	14.4	80.0	110	2	US-08-431-333-11	Sequence 11, Appl
c	42	14.4	80.0	110	5	PCT-US91-02321-11	Sequence 11, Appl
c	43	14.4	80.0	306	2	US-08-634-797-14	Sequence 14, Appl
	44	14.4	80.0	707	2	US-08-850-910A-40	Sequence 40, Appl
	45	14.4	80.0	971	3	US-09-248-335-65	Sequence 65, Appl

ALIGNMENTS

RESULT 1
US-08-480-917-8
; Sequence 8, Application US/08480917
; Patent No. 5820864
; GENERAL INFORMATION:
; APPLICANT: PARANHOS-BACCALA, Glaucia
; APPLICANT: LESENECHAL, Mylene
; APPLICANT: JOLIVET, Michel
; TITLE OF INVENTION: NEW TRYPANOSOMA CRUZI ANTIGEN, AND GENE
; TITLE OF INVENTION: ENCODING THE LATTER; THEIR APPLICATION TO THE DETECTION OF
; TITLE OF INVENTION: CHAGAS DISEASE
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oliff & Berridge
; STREET: 700 South Washington Street, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,917
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 36400
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-480-917-8

Query Match 100.0%; Score 18; DB 1; Length 18;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgcagcagcggcagaagt 18

```
          ||||||||||||||||||
Db      1 TGCAGCAGCGGCAGAAGT 18


RESULT   2
US-08-480-917-1/c
; Sequence 1, Application US/08480917
; Patent No. 5820864
;  GENERAL INFORMATION:
;    APPLICANT:  PARANHOS-BACCALA, Glaucia
;    APPLICANT:  LESENECHAL, Mylene
;    APPLICANT:  JOLIVET, Michel
;    TITLE OF INVENTION:  NEW TRYPANOSOMA CRUZI ANTIGEN, AND GENE
;    TITLE OF INVENTION:  ENCODING THE LATTER; THEIR APPLICATION TO THE DETECTION OF
;    TITLE OF INVENTION:  CHAGAS DISEASE
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Oliff & Berridge
;      STREET:  700 South Washington Street, Suite 300
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22314
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,917
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berridge, William P.
;      REGISTRATION NUMBER:  30,024
;      REFERENCE/DOCKET NUMBER:  WPB 36400
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-836-6400
;      TELEFAX:  703-836-2787
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3402 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-480-917-1


  Query Match             100.0%;  Score 18;  DB 1;  Length 3402;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||||||||||||||||
Db   1459 TGCAGCAGCGGCAGAAGT 1442


RESULT   3
US-08-970-269A-32
; Sequence 32, Application US/08970269A
; Patent No. 5976803
;  GENERAL INFORMATION:
;    APPLICANT:  Katheryn Meek
;    TITLE OF INVENTION:  Genetic Test For Equine Severe
;    TITLE OF INVENTION:  Combined Immunodeficiency Disease
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dr. Benjamin A. Adler
;      STREET:  8011 Candle Lane
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Apple
;      OPERATING SYSTEM:  Macintosh
;      SOFTWARE:  Microsoft Word for Macintosh
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/970,269A
;      FILING DATE:  No. 5976803ember 14, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Adler Ph.D., Benjamin A.
;      REGISTRATION NUMBER:  35,423
;      REFERENCE/DOCKET NUMBER:  D5860
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  713-777-2321
;      TELEFAX:  713-777-6908
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11873 bp
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:
;      DESCRIPTION:  other nucleic acid
;    HYPOTHETICAL:  no
;    ANTI-SENSE:  no
;    ORIGINAL SOURCE:
;    FEATURE:
US-08-970-269A-32


  Query Match             100.0%;  Score 18;  DB 2;  Length 11873;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||||||||||||||||
Db   6429 TGCAGCAGCGGCAGAAGT 6446


RESULT   4
US-08-970-269A-31
; Sequence 31, Application US/08970269A
; Patent No. 5976803
;  GENERAL INFORMATION:
;    APPLICANT:  Katheryn Meek
;    TITLE OF INVENTION:  Genetic Test For Equine Severe
;    TITLE OF INVENTION:  Combined Immunodeficiency Disease
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dr. Benjamin A. Adler
;      STREET:  8011 Candle Lane
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Apple
;      OPERATING SYSTEM:  Macintosh
;      SOFTWARE:  Microsoft Word for Macintosh
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/970,269A
;      FILING DATE:  No. 5976803ember 14, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Adler Ph.D., Benjamin A.
;      REGISTRATION NUMBER:  35,423
;      REFERENCE/DOCKET NUMBER:  D5860
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  713-777-2321
;      TELEFAX:  713-777-6908
```

```
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11878 bp
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:
;      DESCRIPTION:  other nucleic acid
;    HYPOTHETICAL:  no
;    ANTI-SENSE:  no
;    ORIGINAL SOURCE:
;    FEATURE:
US-08-970-269A-31


  Query Match           100.0%;  Score 18;  DB 2;  Length 11878;
  Best Local Similarity  100.0%;  Pred. No. 12;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||||||||||||||||
Db   6429 TGCAGCAGCGGCAGAAGT 6446


RESULT   5
US-08-970-269A-28
; Sequence 28, Application US/08970269A
; Patent No. 5976803
;  GENERAL INFORMATION:
;    APPLICANT:  Katheryn Meek
;    TITLE OF INVENTION:  Genetic Test For Equine Severe
;    TITLE OF INVENTION:  Combined Immunodeficiency Disease
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dr. Benjamin A. Adler
;      STREET:  8011 Candle Lane
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Apple
;      OPERATING SYSTEM:  Macintosh
;      SOFTWARE:  Microsoft Word for Macintosh
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/970,269A
;      FILING DATE:  No. 5976803ember 14, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Adler Ph.D., Benjamin A.
;      REGISTRATION NUMBER:  35,423
;      REFERENCE/DOCKET NUMBER:  D5860
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  713-777-2321
;      TELEFAX:  713-777-6908
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11883 bp
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:
;      DESCRIPTION:  other nucleic acid
;    HYPOTHETICAL:  no
;    ANTI-SENSE:  no
;    ORIGINAL SOURCE:
US-08-970-269A-28


  Query Match           100.0%;  Score 18;  DB 2;  Length 11883;
  Best Local Similarity  100.0%;  Pred. No. 12;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||||||||||||||||
Db   6432 TGCAGCAGCGGCAGAAGT 6449


RESULT   6
US-08-923-137-2
; Sequence 2, Application US/08923137
; Patent No. 6083716
;  GENERAL INFORMATION:
;    APPLICANT:  Wilson, James M.
;    APPLICANT:  Farina, Steven F.
;    APPLICANT:  Fisher, Krishna J.
;    TITLE OF INVENTION:  Chimpanzee Adenovirus Vectors
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Howson and Howson
;      STREET:  Spring House Corporate Cntr., P.O. Box 457
;      CITY:  Spring House
;      STATE:  Pennsylvania
;      COUNTRY:  United States of America
;      ZIP:  19477
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/923,137
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/024,700
;      FILING DATE:  06-SEP-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bak, Mary E.
;      REGISTRATION NUMBER:  31,215
;      REFERENCE/DOCKET NUMBER:  GNVPN.021CIP1USA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-540-9200
;      TELEFAX:  215-540-5818
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  36519 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
US-08-923-137-2


  Query Match           91.1%;  Score 16.4;  DB 3;  Length 36519;
  Best Local Similarity  94.4%;  Pred. No. 63;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          |||||||||| |||||||
Db  31662 TGCAGCAGCGCCAGAAGT 31679


RESULT   7
US-09-110-959A-13
; Sequence 13, Application US/09110959A
; Patent No. 6268197
; GENERAL INFORMATION:
;  APPLICANT:  Schulein, Martin
;  APPLICANT:  Outtrup, Helle
;  APPLICANT:  Jorgensen, Per Lina
;  APPLICANT:  Bjornvad, Mads Eskelund
;  TITLE OF INVENTION:  Alkaline Xyloglucanase
;  FILE REFERENCE:  5206.200-US
;  CURRENT APPLICATION NUMBER: US/09/110,959A
```

```
;  CURRENT FILING DATE:  1998-07-07
;  PRIOR APPLICATION NUMBER:  0822/97
;  PRIOR FILING DATE:  1997-07-07
;  PRIOR APPLICATION NUMBER:  1213/97
;  PRIOR FILING DATE:  1997-10-24
;  PRIOR APPLICATION NUMBER:  60/054,039
;  PRIOR FILING DATE:  1997-07-28
;  PRIOR APPLICATION NUMBER:  60/063,694
;  PRIOR FILING DATE:  1997-10-28
;  NUMBER OF SEQ ID NOS:  14
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 13
;   LENGTH: 39
;   TYPE: DNA
;   ORGANISM: Bacillus sp.
US-09-110-959A-13


  Query Match             85.6%;  Score 15.4;  DB 4;  Length 39;
  Best Local Similarity   94.1%;  Pred. No. 98;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaag 17
          |||||| ||||||||||
Db      6 tgcagccgcggcagaag 22


RESULT   8
US-09-177-325-2
; Sequence 2, Application US/09177325B
; Patent No. 6214983
; GENERAL INFORMATION:
;  APPLICANT:      Tureci, Ozlem
;  APPLICANT:  Sakin, Ugur
;  APPLICANT:  Pfreundschuh, Michael
;  TITLE OF INVENTION:      Isolated Nucleic Acid Molecules Which Encode SCP Proteins, A
;  TITLE OF INVENTION:  Thereof
;  FILE REFERENCE:      LUD 5525
;  CURRENT APPLICATION NUMBER: US/09/177,325B
;  CURRENT FILING DATE:  1998-10-22
;  NUMBER OF SEQ ID NOS:      15
; SEQ ID NO 2
;   LENGTH: 1276
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
US-09-177-325-2


  Query Match             85.6%;  Score 15.4;  DB 4;  Length 1276;
  Best Local Similarity   94.1%;  Pred. No. 1.3e+02;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 gcagcagcggcagaagt 18
          |||||||| ||||||||
Db    711 gcagcagcagcagaagt 727


RESULT   9
US-07-923-739-1
; Sequence 1, Application US/07923739
; Patent No. 5401835
;  GENERAL INFORMATION:
;    APPLICANT:  Chishti, Athar H.
;    TITLE OF INVENTION:  Human Erythroid p55 and Methods of Use
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Harness, Dickey and Pierce
;      STREET:  5445 Corporate Drive
;      CITY:  Troy
;      STATE:  Michigan
;      COUNTRY:  US
;      ZIP:  48098
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/923,739
;      FILING DATE:  19920731
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lewak, Anna M
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (313) 641-1600
;      TELEFAX:  (313) 641-0270
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2000 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  115..1512
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /product= "protein"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /note= "Polyadenylation consensus sequence is at
;      OTHER INFORMATION:  nucleic acid residues 1982-1987.  Amino acid
;      OTHER INFORMATION:  residues 165-233 represent the SH-3 motif."
;    PUBLICATION INFORMATION:
;      AUTHORS:  Andrabi, Khurshid
;      AUTHORS:  Rana, Ajay
;      AUTHORS:  Keeler, Marilyn
;      AUTHORS:  Maalouf, George
;      AUTHORS:  Bruns, Gail
;      AUTHORS:  Chishti, Athar
;      TITLE:  Human erythroid p55: Homolog of Drosophilia
;      TITLE:  tumor suppressor factor is highly conserved
;      TITLE:  X-linked gene product with guanylate kinase
;      TITLE:  activity
;      JOURNAL:  J. Biol. Chem.
;      DATE:  1992
;    PUBLICATION INFORMATION:
;      AUTHORS:  Ruff, Paul
;      AUTHORS:  Speicher, David W.
;      AUTHORS:  Husain-Chishti, A.
;      TITLE:  Molecular identification of a major
;      TITLE:  palmitoylated erythrocyte membrane protein
;      TITLE:  containing the src homology 3 motif
;      JOURNAL:  Proc. Natl. Acad. Sci. U.S.A.
;      VOLUME:  88
;      PAGES:  6595-6599
;      DATE:  August-1991
;      RELEVANT RESIDUES IN SEQ ID NO:  1: FROM 1 TO 2000
;    PUBLICATION INFORMATION:
;      AUTHORS:  Husain-Chishti, Athar
;      AUTHORS:  Faquin, William
;      AUTHORS:  Wu, Chi-Chih
;      AUTHORS:  Branton, Daniel
;      TITLE:  Purification of Erthrocyte of Dematin
;      TITLE:  (Protein 4.9) Reveals an Endogenous Protein Kinase
;      TITLE:  That Modulates Actin-bundling Activity
;      JOURNAL:  J. Biol. Chem.
;      VOLUME:  264
;      ISSUE:  15
;      PAGES:  8985-8991
;      DATE:  5-25-1989
US-07-923-739-1


  Query Match             85.6%;  Score 15.4;  DB 1;  Length 2000;
  Best Local Similarity   94.1%;  Pred. No. 1.3e+02;
```

```
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaag 17
          ||||||||| |||||||
Db   1349 TGCAGCAGCTGCAGAAG 1365


RESULT  10
US-08-864-038A-1/c
; Sequence 1, Application US/08864038A
; Patent No. 6001592
;  GENERAL INFORMATION:
;    APPLICANT:  Kunio NAKASHIMA et al.
;    TITLE OF INVENTION:  NOVEL POLYPEPTIDE GENE cDNA, VECTOR
;    TITLE OF INVENTION:  CONTAINING SAID cDNA, HOST CELLS TRANSFORMED WITH SAID
;    TITLE OF INVENTION:  VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
;    TITLE OF INVENTION:  SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY
;    TITLE OF INVENTION:  TO SAID POLYPEPTIDE
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  812-5 Hirano
;      STREET:  Isshinden
;      CITY:  Tsu-city
;      STATE:  Mie-prefecture
;      COUNTRY:  JAPAN
;      ZIP:  514-01
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.44 MB storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Microsoft Windows 95
;      SOFTWARE:  Word Perfect 6.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/864,038A
;      FILING DATE:  May 28, 1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 8-184459
;      FILING DATE:  15-July-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  C. Bruce Hamburg
;      REGISTRATION NUMBER:  22,389
;      REFERENCE/DOCKET NUMBER:  F-5610
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)986-2340
;      TELEFAX:  (212)953-7733
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2214
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    ORIGINAL SOURCE:
;      ORGANISM:  Pinctada fucata
;      CELL TYPE:  mantle epithelial cell
US-08-864-038A-1


  Query Match             85.6%;  Score 15.4;  DB 3;  Length 2214;
  Best Local Similarity   94.1%;  Pred. No. 1.4e+02;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaag 17
          |||||||||||||| ||
Db   1767 TGCAGCAGCGGCAGCAG 1751


RESULT  11
US-08-864-038A-2/c
; Sequence 2, Application US/08864038A
; Patent No. 6001592
;  GENERAL INFORMATION:
;    APPLICANT:  Kunio NAKASHIMA et al.
;    TITLE OF INVENTION:  NOVEL POLYPEPTIDE GENE cDNA, VECTOR
;    TITLE OF INVENTION:  CONTAINING SAID cDNA, HOST CELLS TRANSFORMED WITH SAID
;    TITLE OF INVENTION:  VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
;    TITLE OF INVENTION:  SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBOD
;    TITLE OF INVENTION:  TO SAID POLYPEPTIDE
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  812-5 Hirano
;      STREET:  Isshinden
;      CITY:  Tsu-city
;      STATE:  Mie-prefecture
;      COUNTRY:  JAPAN
;      ZIP:  514-01
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.44 MB storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Microsoft Windows 95
;      SOFTWARE:  Word Perfect 6.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/864,038A
;      FILING DATE:  May 28, 1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 8-184459
;      FILING DATE:  15-July-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  C. Bruce Hamburg
;      REGISTRATION NUMBER:  22,389
;      REFERENCE/DOCKET NUMBER:  F-5610
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)986-2340
;      TELEFAX:  (212)953-7733
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3331
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Pinctada fucata
;      CELL TYPE:  mantle epithelial cell
;    FEATURE:  mRNA
;      LOCATION:  from 1 to 3331
;      IDENTIFICATION METHOD:  E (by experiment)
US-08-864-038A-2


  Query Match             85.6%;  Score 15.4;  DB 3;  Length 3331;
  Best Local Similarity   94.1%;  Pred. No. 1.4e+02;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaag 17
          |||||||||||||| ||
Db   1816 TGCAGCAGCGGCAGCAG 1800


RESULT  12
US-08-864-038A-4/c
; Sequence 4, Application US/08864038A
; Patent No. 6001592
;  GENERAL INFORMATION:
;    APPLICANT:  Kunio NAKASHIMA et al.
;    TITLE OF INVENTION:  NOVEL POLYPEPTIDE GENE cDNA, VECTOR
;    TITLE OF INVENTION:  CONTAINING SAID cDNA, HOST CELLS TRANSFORMED WITH SAID
;    TITLE OF INVENTION:  VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
;    TITLE OF INVENTION:  SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBOD
;    TITLE OF INVENTION:  TO SAID POLYPEPTIDE
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  812-5 Hirano
;      STREET:  Isshinden
;      CITY:  Tsu-city
;      STATE:  Mie-prefecture
;      COUNTRY:  JAPAN
;      ZIP:  514-01
```

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.44 MB storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Microsoft Windows 95
;      SOFTWARE:  Word Perfect 6.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/864,038A
;      FILING DATE:  May 28, 1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 8-184459
;      FILING DATE:  15-July-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  C. Bruce Hamburg
;      REGISTRATION NUMBER:  22,389
;      REFERENCE/DOCKET NUMBER:  F-5610
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)986-2340
;      TELEFAX:  (212)953-7733
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3331
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Pinctada fucata
;      CELL TYPE:  mantle epithelial cell
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  from 50 to 2263
;      IDENTIFICATION METHOD:  P (by similarity to some other pattern)
US-08-864-038A-4


  Query Match             85.6%;  Score 15.4;  DB 3;  Length 3331;
  Best Local Similarity   94.1%;  Pred. No. 1.4e+02;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 tgcagcagcggcagaag 17
           ||||||||||||||| ||
Db    1816 TGCAGCAGCGGCAGCAG 1800


RESULT  13
US-08-131-365B-53
; Sequence 53, Application US/08131365B
; Patent No. 5527690
;  GENERAL INFORMATION:
;    APPLICANT:  Brown, Michael S.
;    APPLICANT:  Briggs, Michael R.
;    APPLICANT:  Wang, Xiaodong
;    APPLICANT:  Goldstein, Joseph L.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS RELATING
;    TITLE OF INVENTION:  TO STEROL REGULATORY ELEMENT BINDING
;    TITLE OF INVENTION:  PROTEINS
;    NUMBER OF SEQUENCES:  64
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  U.S.A.
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/131,365B
;      FILING DATE:  01-OCT-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, David L.
;      REGISTRATION NUMBER:  32,165
;      REFERENCE/DOCKET NUMBER:  UTSD:372/PAR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  53:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5197 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "DNA"
US-08-131-365B-53


  Query Match             85.6%;  Score 15.4;  DB 1;  Length 5197;
  Best Local Similarity   94.1%;  Pred. No. 1.5e+02;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 gcagcagcggcagaagt 18
           ||||||||||||||| |||
Db     302 GCAGCAGCGGCAGCAGT 318


RESULT  14
US-08-668-123-53
; Sequence 53, Application US/08668123
; Patent No. 5891631
;  GENERAL INFORMATION:
;    APPLICANT:  Brown, Michael S.
;    APPLICANT:  Briggs, Michael R.
;    APPLICANT:  Wang, Xiaodong
;    APPLICANT:  Goldstein, Joseph L.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS RELATING
;    TITLE OF INVENTION:  TO STEROL REGULATORY ELEMENT BINDING
;    TITLE OF INVENTION:  PROTEINS
;    NUMBER OF SEQUENCES:  64
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  U.S.A.
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/668,123
;      FILING DATE:  14-JUN-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/131,365
;      FILING DATE:  01-OCT-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, David L.
;      REGISTRATION NUMBER:  32,165
;      REFERENCE/DOCKET NUMBER:  UTSD:372/PAR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  53:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5197 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
```

```
;      DESCRIPTION:  /desc = "DNA"
US-08-668-123-53


  Query Match             85.6%;  Score 15.4;  DB 2;  Length 5197;
  Best Local Similarity   94.1%;  Pred. No. 1.5e+02;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 gcagcagcggcagaagt 18
           |||||||||||||| |||
Db     302 GCAGCAGCGGCAGCAGT 318


RESULT  15
US-08-934-494-1
; Sequence 1, Application US/08934494
; Patent No. 6030831
;  GENERAL INFORMATION:
;    APPLICANT:  Gurney, Austin
;    TITLE OF INVENTION:  Tie Ligands
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/934,494
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dreger, Ginger R.
;      REGISTRATION NUMBER:  33,055
;      REFERENCE/DOCKET NUMBER:  P1078
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-3216
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1869 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-934-494-1


  Query Match             83.3%;  Score 15;  DB 3;  Length 1869;
  Best Local Similarity   100.0%;  Pred. No. 2e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 cagcagcggcagaag 17
           |||||||||||||||
Db    1268 CAGCAGCGGCAGAAG 1282


Search completed: September 21, 2001, 02:16:34
Job time: 28227 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: September 21, 2001, 02:19:42 ; Search time 6407.28 Seconds
(without alignments)
41.513 Million cell updates/sec

Title: US-09-138-735-8
Perfect score: 18
Sequence: 1 tgcagcagcggcagaagt 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 14155048 seqs, 7388405095 residues

Total number of hits satisfying chosen parameters: 28310096

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*
1: /cgn1_7/ptodata/1/pna/PCTUS_COMB.seq:*
2: /cgn1_7/ptodata/1/pna/US06_COMB.seq:*
3: /cgn1_7/ptodata/1/pna/US07_COMB.seq:*
4: /cgn1_7/ptodata/1/pna/US080_COMB.seq:*
5: /cgn1_7/ptodata/1/pna/US081_COMB.seq:*
6: /cgn1_7/ptodata/1/pna/US082_COMB.seq:*
7: /cgn1_7/ptodata/1/pna/US083_COMB.seq:*
8: /cgn1_7/ptodata/1/pna/US084_COMB.seq:*
9: /cgn1_7/ptodata/1/pna/US085_COMB.seq:*
10: /cgn1_7/ptodata/1/pna/US086_COMB.seq:*
11: /cgn1_7/ptodata/1/pna/US087_COMB.seq:*
12: /cgn1_7/ptodata/1/pna/US088_COMB.seq:*
13: /cgn1_7/ptodata/1/pna/US089_COMB.seq:*
14: /cgn1_7/ptodata/1/pna/US090_COMB.seq:*
15: /cgn1_7/ptodata/1/pna/US091_COMB.seq:*
16: /cgn1_7/ptodata/1/pna/US092_COMB.seq:*
17: /cgn1_7/ptodata/1/pna/US093_COMB.seq:*
18: /cgn1_7/ptodata/1/pna/US094_COMB.seq:*
19: /cgn1_7/ptodata/1/pna/US095A_COMB.seq:*
20: /cgn1_7/ptodata/1/pna/US095B_COMB.seq:*
21: /cgn1_7/ptodata/1/pna/US095C_COMB.seq:*
22: /cgn1_7/ptodata/1/pna/US095D_COMB.seq:*
23: /cgn1_7/ptodata/1/pna/US096A_COMB.seq:*
24: /cgn1_7/ptodata/1/pna/US096B_COMB.seq:*
25: /cgn1_7/ptodata/1/pna/US096C_COMB.seq:*
26: /cgn1_7/ptodata/1/pna/US096D_COMB.seq:*
27: /cgn1_7/ptodata/1/pna/US096E_COMB.seq:*
28: /cgn1_7/ptodata/1/pna/US097A_COMB.seq:*
29: /cgn1_7/ptodata/1/pna/US097B_COMB.seq:*
30: /cgn1_7/ptodata/1/pna/US097C_COMB.seq:*
31: /cgn1_7/ptodata/1/pna/US098_COMB.seq:*
32: /cgn1_7/ptodata/1/pna/US6000_COMB.seq:*
33: /cgn1_7/ptodata/1/pna/US6001_COMB.seq:*
34: /cgn1_7/ptodata/1/pna/US6002_COMB.seq:*
35: /cgn1_7/ptodata/1/pna/US6003_COMB.seq:*
36: /cgn1_7/ptodata/1/pna/US6004_COMB.seq:*
37: /cgn1_7/ptodata/1/pna/US6005_COMB.seq:*
38: /cgn1_7/ptodata/1/pna/US6006_COMB.seq:*
39: /cgn1_7/ptodata/1/pna/US6007_COMB.seq:*
40: /cgn1_7/ptodata/1/pna/US6008_COMB.seq:*
41: /cgn1_7/ptodata/1/pna/US6009_COMB.seq:*
42: /cgn1_7/ptodata/1/pna/US6010_COMB.seq:*
43: /cgn1_7/ptodata/1/pna/US6011_COMB.seq:*
44: /cgn1_7/ptodata/1/pna/US6012_COMB.seq:*
45: /cgn1_7/ptodata/1/pna/US6013_COMB.seq:*
46: /cgn1_7/ptodata/1/pna/US6014_COMB.seq:*
47: /cgn1_7/ptodata/1/pna/US6015_COMB.seq:*
48: /cgn1_7/ptodata/1/pna/US6016_COMB.seq:*
49: /cgn1_7/ptodata/1/pna/US6017_COMB.seq:*
50: /cgn1_7/ptodata/1/pna/US6018_COMB.seq:*
51: /cgn1_7/ptodata/1/pna/US6019_COMB.seq:*
52: /cgn1_7/ptodata/1/pna/US6020_COMB.seq:*
53: /cgn1_7/ptodata/1/pna/US6021_COMB.seq:*
54: /cgn1_7/ptodata/1/pna/US6022_COMB.seq:*
55: /cgn1_7/ptodata/1/pna/US6023_COMB.seq:*
56: /cgn1_7/ptodata/1/pna/US6024_COMB.seq:*
57: /cgn1_7/ptodata/1/pna/US6025_COMB.seq:*
58: /cgn1_7/ptodata/1/pna/US6026_COMB.seq:*
59: /cgn1_7/ptodata/1/pna/US6027_COMB.seq:*
60: /cgn1_7/ptodata/1/pna/US6028_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	18	100.0	18	13	US-08-988-242-4	Sequence 4, Appli
	2	18	100.0	18	15	US-09-138-736-8	Sequence 8, Appli
	3	18	100.0	328	1	PCT-US01-01329-161	Sequence 161, App
	4	18	100.0	328	1	PCT-US01-01339-1027	Sequence 1027, Ap
	5	18	100.0	444	30	US-09-770-175-4591	Sequence 4591, Ap
	6	18	100.0	475	16	US-09-240-371-7431	Sequence 7431, Ap
	7	18	100.0	475	16	US-09-293-972-5995	Sequence 5995, Ap
	8	18	100.0	484	23	US-09-606-776-3048	Sequence 3048, Ap
	9	18	100.0	484	25	US-09-644-867-3664	Sequence 3664, Ap
	10	18	100.0	506	6	US-08-276-163A-13709	Sequence 13709, A
	11	18	100.0	506	6	US-08-276-163B-13709	Sequence 13709, A
	12	18	100.0	506	6	US-08-276-163D-13709	Sequence 13709, A
	13	18	100.0	508	25	US-09-652-914-2822	Sequence 2822, Ap
	14	18	100.0	940	23	US-09-606-776-5253	Sequence 5253, Ap
	15	18	100.0	940	30	US-09-770-175-8299	Sequence 8299, Ap
	16	18	100.0	1589	25	US-09-641-377-783	Sequence 783, App
c	17	18	100.0	1707	25	US-09-644-867-5948	Sequence 5948, Ap
c	18	18	100.0	1707	25	US-09-652-914-7995	Sequence 7995, Ap
c	19	18	100.0	3402	13	US-08-988-242-1	Sequence 1, Appli
c	20	18	100.0	3402	15	US-09-138-736-1	Sequence 1, Appli
	21	18	100.0	11873	18	US-09-407-562-32	Sequence 32, Appl
	22	18	100.0	11878	18	US-09-407-562-31	Sequence 31, Appl
	23	18	100.0	11883	18	US-09-407-562-28	Sequence 28, Appl
	24	17	94.4	429	49	US-60-172-362-136	Sequence 136, App
c	25	17	94.4	718	25	US-09-654-617-126156	Sequence 126156,
c	26	17	94.4	718	27	US-09-684-016-126156	Sequence 126156,
	27	17	94.4	2121	18	US-09-489-039A-5009	Sequence 5009, Ap
	28	17	94.4	91470	20	US-09-534-859-19	Sequence 19, Appl
c	29	16.4	91.1	249	18	US-09-436-762A-26553	Sequence 26553, A
c	30	16.4	91.1	280	15	US-09-127-612-1967	Sequence 1967, Ap
c	31	16.4	91.1	280	21	US-09-540-208-69979	Sequence 69979, A
c	32	16.4	91.1	281	15	US-09-127-612-2550	Sequence 2550, Ap
c	33	16.4	91.1	281	21	US-09-540-208-68661	Sequence 68661, A
	34	16.4	91.1	284	48	US-60-164-443-401	Sequence 401, App
	35	16.4	91.1	307	20	US-09-535-897-16535	Sequence 16535, A
c	36	16.4	91.1	339	21	US-09-540-229-178511	Sequence 178511,
c	37	16.4	91.1	402	17	US-09-333-535A-2478	Sequence 2478, Ap
c	38	16.4	91.1	402	25	US-09-654-617-56139	Sequence 56139, A
c	39	16.4	91.1	402	27	US-09-684-016-56139	Sequence 56139, A
c	40	16.4	91.1	412	56	US-60-248-798-473	Sequence 473, App
c	41	16.4	91.1	423	25	US-09-654-617-258006	Sequence 258006,
c	42	16.4	91.1	423	27	US-09-684-016-258006	Sequence 258006,
	43	16.4	91.1	458	18	US-09-401-645-3287	Sequence 3287, Ap
c	44	16.4	91.1	486	25	US-09-644-870-269	Sequence 269, App
c	45	16.4	91.1	489	53	US-60-217-080-20722	Sequence 20722, A

ALIGNMENTS

```
RESULT   1
US-08-988-242-4
; Sequence 4, Application US/08988242
;  GENERAL INFORMATION:
;    APPLICANT:  PARANHOS-BACCALA, GLAUCIA
;    APPLICANT:  LESENECHAL, MYLENE
;    APPLICANT:  JOLIVET, MICHEL
;    APPLICANT:  MANDRAND, BERNARD
;    TITLE OF INVENTION:  NEW TRYPANOSOMA CRUZI ANTIGEN, GENE
;    TITLE OF INVENTION:  ENCODING THEREFOR, AND METHODS OF DETECTING AND TREATING
;    TITLE OF INVENTION:  CHAGAS DISEASE
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OLIFF & BERRIDGE, PLC
;      STREET:  P.O. BOX 19928
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22320
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/988,242
;      FILING DATE:  10-DEC-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berridge, William P.
;      REGISTRATION NUMBER:  30,024
;      REFERENCE/DOCKET NUMBER:  WPB 36400A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-836-6400
;      TELEFAX:  703-836-2787
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  18 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-988-242-4


  Query Match             100.0%;  Score 18;  DB 13;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 7.2e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||||||||||||||||
Db      1 TGCAGCAGCGGCAGAAGT 18


RESULT   2
US-09-138-736-8
; Sequence 8, Application US/09138736
;  GENERAL INFORMATION:
;    APPLICANT:  PARANHOS-BACCALA, Glaucia
;    APPLICANT:  LESENECHAL, Mylene
;    APPLICANT:  JOLIVET, Michel
;    TITLE OF INVENTION:  NEW TRYPANOSOMA CRUZI ANTIGEN, AND GENE
;    TITLE OF INVENTION:  ENCODING THE LATTER; THEIR APPLICATION TO THE DETECTION OF
;    TITLE OF INVENTION:  CHAGAS DISEASE
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Oliff & Berridge
;      STREET:  700 South Washington Street, Suite 300
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22314
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/138,736
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/480,917
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berridge, William P.
;      REGISTRATION NUMBER:  30,024
;      REFERENCE/DOCKET NUMBER:  WPB 36400
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-836-6400
;      TELEFAX:  703-836-2787
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  18 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-138-736-8


  Query Match             100.0%;  Score 18;  DB 15;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 7.2e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||||||||||||||||
Db      1 TGCAGCAGCGGCAGAAGT 18


RESULT   3
PCT-US01-01329-161
; Sequence 161, Application PC/TUS0101329
; GENERAL INFORMATION:
;  APPLICANT: Human Genome Sciences, Inc., et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA120PCT
;  CURRENT APPLICATION NUMBER: PCT/US01/01329
;  CURRENT FILING DATE:  2001-01-17
;  NUMBER OF SEQ ID NOS: 3506
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 161
;    LENGTH: 328
;    TYPE: DNA
;    ORGANISM: Homo sapiens
PCT-US01-01329-161


  Query Match             100.0%;  Score 18;  DB 1;  Length 328;
  Best Local Similarity   100.0%;  Pred. No. 7.9e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||||||||||||||||
Db    234 tgcagcagcggcagaagt 251


RESULT   4
PCT-US01-01339-1027
; Sequence 1027, Application PC/TUS0101339
; GENERAL INFORMATION:
;  APPLICANT: Human Genome Sciences, Inc., et al.
```

```
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC006PCT
;  CURRENT APPLICATION NUMBER: PCT/US01/01339
;  CURRENT FILING DATE:  2001-03-17
;  NUMBER OF SEQ ID NOS: 10231
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1027
;    LENGTH: 328
;    TYPE: DNA
;    ORGANISM: Homo sapiens
PCT-US01-01339-1027


  Query Match             100.0%;  Score 18;  DB 1;  Length 328;
  Best Local Similarity   100.0%;  Pred. No. 7.9e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||||||||||||||||
Db    234 tgcagcagcggcagaagt 251


RESULT   5
US-09-770-175-4591
; Sequence 4591, Application US/09770175
; GENERAL INFORMATION:
;  APPLICANT: Gearing, David P.
;  APPLICANT:  Holtzman, Douglas A.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
;  TITLE OF INVENTION:  THEREFOR
;  FILE REFERENCE: 1600.2058-001
;  CURRENT APPLICATION NUMBER: US/09/770,175
;  CURRENT FILING DATE:  2001-01-26
;  PRIOR APPLICATION NUMBER: US 60/178,874
;  PRIOR FILING DATE: 2000-01-28
;  NUMBER OF SEQ ID NOS: 8967
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4591
;    LENGTH: 444
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (1)...(444)
;    OTHER INFORMATION: n = A,T,C or G
US-09-770-175-4591


  Query Match             100.0%;  Score 18;  DB 30;  Length 444;
  Best Local Similarity   100.0%;  Pred. No. 8e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||||||||||||||||
Db    255 tgcagcagcggcagaagt 272


RESULT   6
US-09-240-371-7431
; Sequence 7431, Application US/09240371
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION: FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-761
;  CURRENT APPLICATION NUMBER: US/09/240,371
;  CURRENT FILING DATE: 1999-01-29
;  NUMBER OF SEQ ID NOS: 10214
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 7431
;    LENGTH: 475
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (1)...(475)
;    OTHER INFORMATION: n = A,T,C or G
US-09-240-371-7431


  Query Match             100.0%;  Score 18;  DB 16;  Length 475;
  Best Local Similarity   100.0%;  Pred. No. 8e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||||||||||||||||
Db    315 tgcagcagcggcagaagt 332


RESULT   7
US-09-293-972-5995
; Sequence 5995, Application US/09293972
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION: FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-769
;  CURRENT APPLICATION NUMBER: US/09/293,972
;  CURRENT FILING DATE: 1999-04-15
;  NUMBER OF SEQ ID NOS: 34258
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5995
;    LENGTH: 475
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (1)...(475)
;    OTHER INFORMATION: n = A,T,C or G
US-09-293-972-5995


  Query Match             100.0%;  Score 18;  DB 16;  Length 475;
  Best Local Similarity   100.0%;  Pred. No. 8e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||||||||||||||||
Db    315 tgcagcagcggcagaagt 332


RESULT   8
US-09-606-776-3048
; Sequence 3048, Application US/09606776
; GENERAL INFORMATION:
;  APPLICANT: Holtzman, Douglas A.
;  APPLICANT:  Myers, Paul
;  APPLICANT:  Gearing, David P.
;  APPLICANT:  Pan, Yang
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
;  TITLE OF INVENTION:  THEREFOR
;  FILE REFERENCE: 1600.1129-001
;  CURRENT APPLICATION NUMBER: US/09/606,776
;  CURRENT FILING DATE:  2000-06-27
;  PRIOR APPLICATION NUMBER: 60/141,578
;  PRIOR FILING DATE: 1999-06-29
;  PRIOR APPLICATION NUMBER: 60/141,379
;  PRIOR FILING DATE: 1999-06-28
;  PRIOR APPLICATION NUMBER: 60/141,138
;  PRIOR FILING DATE: 1999-06-28
;  PRIOR APPLICATION NUMBER: 60/141,581
;  PRIOR FILING DATE: 1999-06-29
;  NUMBER OF SEQ ID NOS: 5415
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3048
;    LENGTH: 484
```

```
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-606-776-3048


  Query Match             100.0%;  Score 18;  DB 23;  Length 484;
  Best Local Similarity   100.0%;  Pred. No. 8e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tgcagcagcggcagaagt 18
           ||||||||||||||||||
Db     443 tgcagcagcggcagaagt 465


RESULT   9
US-09-644-867-3664
; Sequence 3664, Application US/09644867
; GENERAL INFORMATION:
;  APPLICANT:  Holtzman, Douglas A.
;  APPLICANT:  Donovan, Michael J.
;  APPLICANT:  Leiby, Kevin R.
;  APPLICANT:  Culpepper, Janice A.
;  TITLE OF INVENTION:  NOVEL NUCLEIC ACID MOLECULES AND USES
;  TITLE OF INVENTION:  THEREFOR
;  FILE REFERENCE:  1600.1173-001
;  CURRENT APPLICATION NUMBER: US/09/644,867
;  CURRENT FILING DATE:  2000-08-28
;  PRIOR APPLICATION NUMBER:  60/151,061
;  PRIOR FILING DATE:  1999-08-27
;  NUMBER OF SEQ ID NOS:  8090
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 3664
;   LENGTH: 484
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-644-867-3664


  Query Match             100.0%;  Score 18;  DB 25;  Length 484;
  Best Local Similarity   100.0%;  Pred. No. 8e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tgcagcagcggcagaagt 18
           ||||||||||||||||||
Db     448 tgcagcagcggcagaagt 465


RESULT  10
US-08-276-163A-13709
; Sequence 13709, Application US/08276163A
;  GENERAL INFORMATION:
;    APPLICANT:  Haseltine, William
;    APPLICANT:  Fitzgerald, Lisa
;    APPLICANT:  Adams, Mark
;    APPLICANT:  Lee, Normal
;    APPLICANT:  Fuldner, Rebecca
;    APPLICANT:  Fleischmann, Robert
;    APPLICANT:  Bult, Carol
;    APPLICANT:  Blake, Judy
;    APPLICANT:  White, Owen
;    APPLICANT:  Clayton, Rebecca
;    APPLICANT:  Pellegrino, Susan
;    TITLE OF INVENTION:  Human Genes, Sequences, and Expression
;    NUMBER OF SEQUENCES:  15312
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Carella, Byrne, Bain, Gilfillan, Cecchi, Stewart, &
;      STREET:  6 Becker Farm Road
;      CITY:  Roseland
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07068
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 5.0
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/276,163A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Olstein, Elliot
;      REGISTRATION NUMBER:  24,025
;      REFERENCE/DOCKET NUMBER:  325800-175
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (201) 994-1700
;      TELEFAX:  (201) 994-1744
;  INFORMATION FOR SEQ ID NO:  13709:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  506 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-276-163A-13709


  Query Match             100.0%;  Score 18;  DB 6;  Length 506;
  Best Local Similarity   100.0%;  Pred. No. 8e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tgcagcagcggcagaagt 18
           ||||||||||||||||||
Db     226 TGCAGCAGCGGCAGAAGT 243


RESULT  11
US-08-276-163B-13709
; Sequence 13709, Application US/08276163B
;  GENERAL INFORMATION:
;    APPLICANT:  Adams, Mark D.
;    APPLICANT:  Bult, Carol J.
;    APPLICANT:  Blake, Judith A.
;    APPLICANT:  Fitzgerald, Lisa
;    APPLICANT:  Fleischmann, Robert D.
;    APPLICANT:  Fraser, Claire M.
;    APPLICANT:  Fuldner, Rebecca A.
;    APPLICANT:  Gocayne, Jeannine D.
;    APPLICANT:  Kelley, Jenny
;    APPLICANT:  Kirkness, Ewen F.
;    APPLICANT:  Sutton, Granger G., III
;    APPLICANT:  Haseltine, William A.
;    APPLICANT:  Rosen, Craig A.
;    APPLICANT:  Ruben, Steve
;    APPLICANT:  Dillon, Patrick J.
;    APPLICANT:  Li, Haodong
;    TITLE OF INVENTION:  Human Genes, Sequences and Expression Products
;    NUMBER OF SEQUENCES:  15314
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/276,163B
;      FILING DATE:  July 15, 1994
;      CLASSIFICATION:  435
```

```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Michele M. Wales
;      REGISTRATION NUMBER:  P-43,975
;      REFERENCE/DOCKET NUMBER:  PO-14
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8439
;  INFORMATION FOR SEQ ID NO:  13709:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  506 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-276-163B-13709


  Query Match           100.0%;  Score 18;  DB 6;  Length 506;
  Best Local Similarity  100.0%;  Pred. No. 8e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||||||||||||||||
Db    226 TGCAGCAGCGGCAGAAGT 243


RESULT 12
US-08-276-163D-13709
; Sequence 13709, Application US/08276163D
; GENERAL INFORMATION:
;  APPLICANT: Adams, et. al.
;  TITLE OF INVENTION: Human Genes, Sequences, and Expression Products
;  FILE REFERENCE: PO14
;  CURRENT APPLICATION NUMBER: US/08/276,163D
;  CURRENT FILING DATE:  1994-07-15
;  NUMBER OF SEQ ID NOS: 15314
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13709
;   LENGTH: 506
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc feature
;   LOCATION: (1)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (11)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (30)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (44)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (46)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (51)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (60)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (66)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (148)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (185)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (306)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (321)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (345)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (359)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (364)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (375)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (385)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (425)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (427)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (428)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (430)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (440)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (443)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (470)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (473)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (476)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (478)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (489)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (492)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (494)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (501)
;   OTHER INFORMATION: n equals a,t,g, or c
US-08-276-163D-13709


  Query Match           100.0%;  Score 18;  DB 6;  Length 506;
  Best Local Similarity  100.0%;  Pred. No. 8e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||||||||||||||||
```

```
Db    226 tgcagcagcggcagaagt 243


RESULT  13
US-09-652-914-2822
; Sequence 2822, Application US/09652914
; GENERAL INFORMATION:
;  APPLICANT:  Shyjan, Andrew W.
;  TITLE OF INVENTION:  NOVEL NUCLEIC ACID MOLECULES AND USES
;  TITLE OF INVENTION:  THEREFOR
;  FILE REFERENCE:  1600.1193-001
;  CURRENT APPLICATION NUMBER: US/09/652,914
;  CURRENT FILING DATE:  2000-08-31
;  PRIOR APPLICATION NUMBER:  60/152,112
;  PRIOR FILING DATE:  1999-08-31
;  NUMBER OF SEQ ID NOS:  9677
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 2822
;   LENGTH: 508
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-652-914-2822


  Query Match            100.0%;  Score 18;  DB 25;  Length 508;
  Best Local Similarity   100.0%;  Pred. No. 8e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||||||||||||||||
Db    219 tgcagcagcggcagaagt 236


RESULT  14
US-09-606-776-5253
; Sequence 5253, Application US/09606776
; GENERAL INFORMATION:
;  APPLICANT: Holtzman, Douglas A.
;  APPLICANT:  Myers, Paul
;  APPLICANT:  Gearing, David P.
;  APPLICANT:  Pan, Yang
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
;  TITLE OF INVENTION:  THEREFOR
;  FILE REFERENCE: 1600.1129-001
;  CURRENT APPLICATION NUMBER: US/09/606,776
;  CURRENT FILING DATE:  2000-06-27
;  PRIOR APPLICATION NUMBER: 60/141,578
;  PRIOR FILING DATE: 1999-06-29
;  PRIOR APPLICATION NUMBER: 60/141,379
;  PRIOR FILING DATE: 1999-06-28
;  PRIOR APPLICATION NUMBER: 60/141,138
;  PRIOR FILING DATE: 1999-06-28
;  PRIOR APPLICATION NUMBER: 60/141,581
;  PRIOR FILING DATE: 1999-06-29
;  NUMBER OF SEQ ID NOS: 5415
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5253
;   LENGTH: 940
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-606-776-5253


  Query Match            100.0%;  Score 18;  DB 23;  Length 940;
  Best Local Similarity   100.0%;  Pred. No. 8.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||||||||||||||||
Db    524 tgcagcagcggcagaagt 541


RESULT  15
US-09-770-175-8299
; Sequence 8299, Application US/09770175
; GENERAL INFORMATION:
;  APPLICANT: Gearing, David P.
;  APPLICANT:  Holtzman, Douglas A.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
;  TITLE OF INVENTION:  THEREFOR
;  FILE REFERENCE: 1600.2058-001
;  CURRENT APPLICATION NUMBER: US/09/770,175
;  CURRENT FILING DATE:  2001-01-26
;  PRIOR APPLICATION NUMBER: US 60/178,874
;  PRIOR FILING DATE: 2000-01-28
;  NUMBER OF SEQ ID NOS: 8967
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8299
;   LENGTH: 940
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-770-175-8299


  Query Match            100.0%;  Score 18;  DB 30;  Length 940;
  Best Local Similarity   100.0%;  Pred. No. 8.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||||||||||||||||
Db    524 tgcagcagcggcagaagt 541


Search completed: September 21, 2001, 02:19:48
Job time: 30460 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: September 21, 2001, 02:39:09 ; Search time 1035.34 Seconds
(without alignments)
34.697 Million cell updates/sec

Title: US-09-138-735-8
Perfect score: 18
Sequence: 1 tgcagcagcggcagaagt 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1741792 seqs, 997871973 residues

Total number of hits satisfying chosen parameters: 3483584

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_New:*
1: /cgn1_7/ptodata/2/pna/PCT_NEW_COMB.seq:*
2: /cgn1_7/ptodata/2/pna/US06_NEW_COMB.seq:*
3: /cgn1_7/ptodata/2/pna/US07_NEW_COMB.seq:*
4: /cgn1_7/ptodata/2/pna/US08_NEW_COMB.seq:*
5: /cgn1_7/ptodata/2/pna/US09_NEW_COMB.seq:*
6: /cgn1_7/ptodata/2/pna/US09_NEW_COMB.seq1:*
7: /cgn1_7/ptodata/2/pna/US09_NEW_COMB.seq2:*
8: /cgn1_7/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	18	100.0	18	6	US-09-138-735-8	Sequence 8, Appli
	2	18	100.0	328	6	US-09-764-891-1027	Sequence 1027, Ap
	3	18	100.0	506	7	US-09-840-145-13709	Sequence 13709, A
c	4	18	100.0	3402	6	US-09-138-735-1	Sequence 1, Appli
	5	17	94.4	91470	6	US-09-803-736-19	Sequence 19, Appl
c	6	16.4	91.1	381	7	US-09-724-671-9810	Sequence 9810, Ap
	7	16.4	91.1	506	6	US-09-804-730-4138	Sequence 4138, Ap
c	8	16.4	91.1	3826	7	US-09-927-091-3	Sequence 3, Appli
c	9	16.4	91.1	30676	7	US-09-927-091-8	Sequence 8, Appli
c	10	16.4	91.1	45845	7	US-09-927-091-6	Sequence 6, Appli
c	11	16.4	91.1	49744	7	US-09-927-091-4	Sequence 4, Appli
c	12	16	88.9	235	7	US-09-724-671-4868	Sequence 4868, Ap
c	13	16	88.9	326	7	US-09-724-671-18201	Sequence 18201, A
c	14	16	88.9	338	7	US-09-724-671-18461	Sequence 18461, A
c	15	16	88.9	363	7	US-09-724-671-18790	Sequence 18790, A
	16	16	88.9	2268	1	PCT-US01-08631-19259	Sequence 19259, A
	17	15.4	85.6	205	7	US-09-867-716-6478	Sequence 6478, Ap
c	18	15.4	85.6	215	8	US-60-253-654-23696	Sequence 23696, A
c	19	15.4	85.6	215	8	US-60-255-592-23696	Sequence 23696, A
	20	15.4	85.6	235	7	US-09-850-147-1383	Sequence 1383, Ap
	21	15.4	85.6	243	6	US-09-790-435-157	Sequence 157, App
c	22	15.4	85.6	256	7	US-09-880-107-541	Sequence 541, App
c	23	15.4	85.6	264	6	US-09-764-864-499	Sequence 499, App
	24	15.4	85.6	266	8	US-60-253-456-735	Sequence 735, App
	25	15.4	85.6	279	8	US-60-253-456-5135	Sequence 5135, Ap
c	26	15.4	85.6	317	8	US-60-253-456-23527	Sequence 23527, A
	27	15.4	85.6	321	7	US-09-724-750-19995	Sequence 19995, A
	28	15.4	85.6	332	6	US-09-758-472-3626	Sequence 3626, Ap
c	29	15.4	85.6	353	8	US-60-252-833-6389	Sequence 6389, Ap
	30	15.4	85.6	359	7	US-09-867-716-17212	Sequence 17212, A
c	31	15.4	85.6	360	7	US-09-864-761-28274	Sequence 26274, A
c	32	15.4	85.6	378	7	US-09-850-118-314	Sequence 314, App
c	33	15.4	85.6	380	5	US-09-847-662-305	Sequence 305, App
c	34	15.4	85.6	389	6	US-09-790-435-2837	Sequence 2837, Ap
c	35	15.4	85.6	391	5	US-09-847-662-1609	Sequence 1609, Ap
	36	15.4	85.6	393	8	US-60-252-833-42024	Sequence 42024, A
c	37	15.4	85.6	421	7	US-09-834-366-3070	Sequence 3070, Ap
	38	15.4	85.6	430	6	US-09-790-435-6901	Sequence 6901, Ap
c	39	15.4	85.6	435	6	US-09-790-435-852	Sequence 852, App
	40	15.4	85.6	440	6	US-09-906-555-12800	Sequence 12800, A
	41	15.4	85.6	441	7	US-09-834-366-38665	Sequence 38665, A
	42	15.4	85.6	449	8	US-60-253-652-17754	Sequence 17754, A
c	43	15.4	85.6	450	7	US-09-864-761-11696	Sequence 11696, A
	44	15.4	85.6	456	6	US-09-909-629-34330	Sequence 34330, A
c	45	15.4	85.6	456	7	US-09-864-761-1296	Sequence 1296, Ap

ALIGNMENTS

```
RESULT 1
US-09-138-735-8
; Sequence 8, Application US/09138735
; GENERAL INFORMATION:
;  APPLICANT: PARANHOS-BACCALA, Glaucia
;  APPLICANT:  LESENECHAL, Mylene
;  APPLICANT:  JOLIVET, Michel
;  TITLE OF INVENTION: TRYPANOSOMA CRUZI ANTIGEN, GENE ENCODING THEREFOR AND METHODS
;  TITLE OF INVENTION:  DETECTING AND TREATING CHAGAS DISEASE
;  FILE REFERENCE: WPB 36400B
;  CURRENT APPLICATION NUMBER: US/09/138,735
;  CURRENT FILING DATE:  1998-08-24
;  PRIOR APPLICATION NUMBER: US 08/480,917
;  PRIOR FILING DATE: 1995-06-07
;  PRIOR APPLICATION NUMBER: FR 94/10132
;  PRIOR FILING DATE: 1994-08-12
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
;   LENGTH: 18
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: probe/primer
US-09-138-735-8


  Query Match             100.0%;  Score 18;  DB 6;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 64;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||||||||||||||||
Db      1 tgcagcagcggcagaagt 18


RESULT 2
US-09-764-891-1027
; Sequence 1027, Application US/09764891
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC006
;  CURRENT APPLICATION NUMBER: US/09/764,891
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 10231
;  SOFTWARE: PatentIn Ver. 2.0
```

```
; SEQ ID NO 1027
;   LENGTH: 328
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-891-1027


  Query Match             100.0%;  Score 18;  DB 6;  Length 328;
  Best Local Similarity   100.0%;  Pred. No. 75;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||||||||||||||||
Db    234 tgcagcagcggcagaagt 251


RESULT   3
US-09-840-145-13709
; Sequence 13709, Application US/09840145
; GENERAL INFORMATION:
;  APPLICANT: Adams, et. al.
;  TITLE OF INVENTION: Human Genes, Sequences, and Expression Products
;  FILE REFERENCE: PO14C1
;  CURRENT APPLICATION NUMBER: US/09/840,145
;  CURRENT FILING DATE:  2001-04-24
;  PRIOR APPLICATION NUMBER: 08/276,163
;  PRIOR FILING DATE: 1994-07-15
;  NUMBER OF SEQ ID NOS: 15314
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13709
;   LENGTH: 506
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (11)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (30)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (44)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (46)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (51)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (60)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (66)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (148)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (185)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (306)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (321)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (345)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (359)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (364)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (375)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (385)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (425)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (427)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (428)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (430)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (440)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (443)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (470)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (473)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (476)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (478)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (489)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (492)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (494)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (501)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-840-145-13709


  Query Match             100.0%;  Score 18;  DB 7;  Length 506;
  Best Local Similarity   100.0%;  Pred. No. 76;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||||||||||||||||
Db    226 tgcagcagcggcagaagt 243


RESULT   4
US-09-138-735-1/c
; Sequence 1, Application US/09138735
; GENERAL INFORMATION:
;  APPLICANT: PARANHOS-BACCALA, Glaucia
;  APPLICANT:  LESENECHAL, Mylene
;  APPLICANT:  JOLIVET, Michel
```

```
;  TITLE OF INVENTION: TRYPANOSOMA CRUZI ANTIGEN, GENE ENCODING THEREFOR AND METHODS OF
;  TITLE OF INVENTION:  DETECTING AND TREATING CHAGAS DISEASE
;  FILE REFERENCE: WPB 36400B
;  CURRENT APPLICATION NUMBER: US/09/138,735
;  CURRENT FILING DATE:  1998-08-24
;  PRIOR APPLICATION NUMBER: US 08/480,917
;  PRIOR FILING DATE: 1995-06-07
;  PRIOR APPLICATION NUMBER: FR 94/10132
;  PRIOR FILING DATE: 1994-08-12
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 3402
;   TYPE: DNA
;   ORGANISM: Trypanosoma cruzi
US-09-138-735-1


  Query Match             100.0%;  Score 18;  DB 6;  Length 3402;
  Best Local Similarity   100.0%;  Pred. No. 85;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||||||||||||||||
Db   1459 TGCAGCAGCGGCAGAAGT 1442


RESULT   5
US-09-803-736-19
; Sequence 19, Application US/09803736
; GENERAL INFORMATION:
;  APPLICANT: Bush, David F.
;  APPLICANT:  Levin, Irena M.
;  APPLICANT:  Norris, Susan R.
;  APPLICANT:  Rounsley, Steven D.
;  APPLICANT:  Wiegand, Roger C.
;  TITLE OF INVENTION: Plant Polymorphic Markers and Uses Thereof
;  FILE REFERENCE: 38-10(15493)D
;  CURRENT APPLICATION NUMBER: US/09/803,736
;  CURRENT FILING DATE:  2001-03-12
;  PRIOR APPLICATION NUMBER: US 09/534,859
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: identified by Attorney Docket number 04983.0206CPUS01 38-10
;  PRIOR FILING DATE: 2000-10-20
;  NUMBER OF SEQ ID NOS: 1582
; SEQ ID NO 19
;   LENGTH: 91470
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-803-736-19


  Query Match             94.4%;  Score 17;  DB 6;  Length 91470;
  Best Local Similarity   100.0%;  Pred. No. 2.8e+02;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaag 17
          |||||||||||||||||
Db  42731 tgcagcagcggcagaag 42747


RESULT   6
US-09-724-671-9810/c
; Sequence 9810, Application US/09724671
; GENERAL INFORMATION:
;  APPLICANT: Watson, James D
;  APPLICANT:  Murison, James G
;  TITLE OF INVENTION: Polynucleotides, polypeptides expressed
;  TITLE OF INVENTION:  by the polynucleotides and methods for their use.
;  FILE REFERENCE: 1050U2
;  CURRENT APPLICATION NUMBER: US/09/724,671
;  CURRENT FILING DATE:  2000-11-28
;  PRIOR APPLICATION NUMBER: U.S. No. 60/171,678
;  PRIOR FILING DATE: 1999-12-23
;  NUMBER OF SEQ ID NOS: 21907
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9810
;   LENGTH: 381
;   TYPE: DNA
;   ORGANISM: Mouse
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(381)
;   OTHER INFORMATION: n = A,T,C or G
US-09-724-671-9810


  Query Match             91.1%;  Score 16.4;  DB 7;  Length 381;
  Best Local Similarity   94.4%;  Pred. No. 3.8e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||| ||||||||||||
Db    106 TGCAGTAGCGGCAGAAGT 89


RESULT   7
US-09-804-730-4138
; Sequence 4138, Application US/09804730
; GENERAL INFORMATION:
;  APPLICANT: Andersen, Scott E.
;  APPLICANT:  Hammond-Kosack, Kim
;  APPLICANT:  Masucci, James D.
;  APPLICANT:  Urban, Martin
;  TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(51838)B
;  CURRENT APPLICATION NUMBER: US/09/804,730
;  CURRENT FILING DATE:  2001-03-13
;  PRIOR APPLICATION NUMBER: US 60/189,657
;  PRIOR FILING DATE: 2000-03-15
;  NUMBER OF SEQ ID NOS: 22828
; SEQ ID NO 4138
;   LENGTH: 506
;   TYPE: DNA
;   ORGANISM: Triticum aestivum
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(506)
;   OTHER INFORMATION: unsure at all n locations
;   OTHER INFORMATION: Clone ID: LIB3350-063-P1-K1-A3
US-09-804-730-4138


  Query Match             91.1%;  Score 16.4;  DB 6;  Length 506;
  Best Local Similarity   94.4%;  Pred. No. 3.8e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||||||| ||||||||
Db    385 tgcagcagccgcagaagt 402


RESULT   8
US-09-927-091-3/c
; Sequence 3, Application US/09927091
; GENERAL INFORMATION:
;  APPLICANT: KILLARY, ANN
;  APPLICANT:  LOTT, STEVE
;  APPLICANT:  CHANDLER, DAWN
;  TITLE OF INVENTION: THE TUMOR SUPPRESSOR CAR-1
;  FILE REFERENCE: UTSC:651US
;  CURRENT APPLICATION NUMBER: US/09/927,091
;  CURRENT FILING DATE:  2001-08-09
;  PRIOR APPLICATION NUMBER: 60/227,560
;  PRIOR FILING DATE: 2000-08-23
```

```
;  PRIOR APPLICATION NUMBER: 60/225,033
;  PRIOR FILING DATE: 2000-08-10
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;    LENGTH: 3826
;    TYPE: DNA
;    ORGANISM: Human
US-09-927-091-3


  Query Match             91.1%;  Score 16.4;  DB 7;  Length 3826;
  Best Local Similarity   94.4%;  Pred. No. 4.3e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 tgcagcagcggcagaagt 18
           |||||| ||||||||||||
Db     706 TGCAGCGGCGGCAGAAGT 689


RESULT   9
US-09-927-091-8/c
; Sequence 8, Application US/09927091
; GENERAL INFORMATION:
;  APPLICANT: KILLARY, ANN
;  APPLICANT:  LOTT, STEVE
;  APPLICANT:  CHANDLER, DAWN
;  TITLE OF INVENTION: THE TUMOR SUPPRESSOR CAR-1
;  FILE REFERENCE: UTSC:651US
;  CURRENT APPLICATION NUMBER: US/09/927,091
;  CURRENT FILING DATE:  2001-08-09
;  PRIOR APPLICATION NUMBER: 60/227,560
;  PRIOR FILING DATE: 2000-08-23
;  PRIOR APPLICATION NUMBER: 60/225,033
;  PRIOR FILING DATE: 2000-08-10
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
;    LENGTH: 30676
;    TYPE: DNA
;    ORGANISM: Human
;    FEATURE:
;    NAME/KEY: modified_base
;    LOCATION: (6671)..(30676)
;    OTHER INFORMATION: n = A or C or G or T/U
US-09-927-091-8


  Query Match             91.1%;  Score 16.4;  DB 7;  Length 30676;
  Best Local Similarity   94.4%;  Pred. No. 4.8e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 tgcagcagcggcagaagt 18
           |||||| ||||||||||||
Db     288 TGCAGCGGCGGCAGAAGT 271


RESULT  10
US-09-927-091-6/c
; Sequence 6, Application US/09927091
; GENERAL INFORMATION:
;  APPLICANT: KILLARY, ANN
;  APPLICANT:  LOTT, STEVE
;  APPLICANT:  CHANDLER, DAWN
;  TITLE OF INVENTION: THE TUMOR SUPPRESSOR CAR-1
;  FILE REFERENCE: UTSC:651US
;  CURRENT APPLICATION NUMBER: US/09/927,091
;  CURRENT FILING DATE:  2001-08-09
;  PRIOR APPLICATION NUMBER: 60/227,560
;  PRIOR FILING DATE: 2000-08-23
;  PRIOR APPLICATION NUMBER: 60/225,033
;  PRIOR FILING DATE: 2000-08-10
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;    LENGTH: 45845
;    TYPE: DNA
;    ORGANISM: Human
US-09-927-091-6


  Query Match             91.1%;  Score 16.4;  DB 7;  Length 45845;
  Best Local Similarity   94.4%;  Pred. No. 4.9e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 tgcagcagcggcagaagt 18
           |||||| ||||||||||||
Db   24868 TGCAGCGGCGGCAGAAGT 24851


RESULT  11
US-09-927-091-4/c
; Sequence 4, Application US/09927091
; GENERAL INFORMATION:
;  APPLICANT: KILLARY, ANN
;  APPLICANT:  LOTT, STEVE
;  APPLICANT:  CHANDLER, DAWN
;  TITLE OF INVENTION: THE TUMOR SUPPRESSOR CAR-1
;  FILE REFERENCE: UTSC:651US
;  CURRENT APPLICATION NUMBER: US/09/927,091
;  CURRENT FILING DATE:  2001-08-09
;  PRIOR APPLICATION NUMBER: 60/227,560
;  PRIOR FILING DATE: 2000-08-23
;  PRIOR APPLICATION NUMBER: 60/225,033
;  PRIOR FILING DATE: 2000-08-10
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;    LENGTH: 49744
;    TYPE: DNA
;    ORGANISM: Human
US-09-927-091-4


  Query Match             91.1%;  Score 16.4;  DB 7;  Length 49744;
  Best Local Similarity   94.4%;  Pred. No. 4.9e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 tgcagcagcggcagaagt 18
           |||||| ||||||||||||
Db   35016 TGCAGCGGCGGCAGAAGT 34999


RESULT  12
US-09-724-671-4868/c
; Sequence 4868, Application US/09724671
; GENERAL INFORMATION:
;  APPLICANT: Watson, James D
;  APPLICANT:  Murison, James G
;  TITLE OF INVENTION: Polynucleotides, polypeptides expressed
;  TITLE OF INVENTION:  by the polynucleotides and methods for their use.
;  FILE REFERENCE: 1050U2
;  CURRENT APPLICATION NUMBER: US/09/724,671
;  CURRENT FILING DATE:  2000-11-28
;  PRIOR APPLICATION NUMBER: U.S. No. 60/171,678
;  PRIOR FILING DATE: 1999-12-23
;  NUMBER OF SEQ ID NOS: 21907
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4868
;    LENGTH: 235
;    TYPE: DNA
;    ORGANISM: Mouse
US-09-724-671-4868


  Query Match             88.9%;  Score 16;  DB 7;  Length 235;
```

```
  Best Local Similarity   100.0%;  Pred. No. 5.5e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 gcagcagcggcagaag 17
          ||||||||||||||||
Db     96 GCAGCAGCGGCAGAAG 81


RESULT  13
US-09-724-671-18201/c
; Sequence 18201, Application US/09724671
; GENERAL INFORMATION:
;  APPLICANT: Watson, James D
;  APPLICANT:  Murison, James G
;  TITLE OF INVENTION: Polynucleotides, polypeptides expressed
;  TITLE OF INVENTION:  by the polynucleotides and methods for their use.
;  FILE REFERENCE: 1050U2
;  CURRENT APPLICATION NUMBER: US/09/724,671
;  CURRENT FILING DATE:  2000-11-28
;  PRIOR APPLICATION NUMBER: U.S. No. 60/171,678
;  PRIOR FILING DATE: 1999-12-23
;  NUMBER OF SEQ ID NOS: 21907
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18201
;   LENGTH: 326
;   TYPE: DNA
;   ORGANISM: Mouse
US-09-724-671-18201


  Query Match             88.9%;  Score 16;  DB 7;  Length 326;
  Best Local Similarity   100.0%;  Pred. No. 5.6e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 gcagcagcggcagaag 17
          ||||||||||||||||
Db     89 GCAGCAGCGGCAGAAG 74


RESULT  14
US-09-724-671-18461/c
; Sequence 18461, Application US/09724671
; GENERAL INFORMATION:
;  APPLICANT: Watson, James D
;  APPLICANT:  Murison, James G
;  TITLE OF INVENTION: Polynucleotides, polypeptides expressed
;  TITLE OF INVENTION:  by the polynucleotides and methods for their use.
;  FILE REFERENCE: 1050U2
;  CURRENT APPLICATION NUMBER: US/09/724,671
;  CURRENT FILING DATE:  2000-11-28
;  PRIOR APPLICATION NUMBER: U.S. No. 60/171,678
;  PRIOR FILING DATE: 1999-12-23
;  NUMBER OF SEQ ID NOS: 21907
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18461
;   LENGTH: 338
;   TYPE: DNA
;   ORGANISM: Mouse
US-09-724-671-18461


  Query Match             88.9%;  Score 16;  DB 7;  Length 338;
  Best Local Similarity   100.0%;  Pred. No. 5.6e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 gcagcagcggcagaag 17
          ||||||||||||||||
Db     89 GCAGCAGCGGCAGAAG 74


RESULT  15
US-09-724-671-18790/c
; Sequence 18790, Application US/09724671
; GENERAL INFORMATION:
;  APPLICANT: Watson, James D
;  APPLICANT:  Murison, James G
;  TITLE OF INVENTION: Polynucleotides, polypeptides expressed
;  TITLE OF INVENTION:  by the polynucleotides and methods for their use.
;  FILE REFERENCE: 1050U2
;  CURRENT APPLICATION NUMBER: US/09/724,671
;  CURRENT FILING DATE:  2000-11-28
;  PRIOR APPLICATION NUMBER: U.S. No. 60/171,678
;  PRIOR FILING DATE: 1999-12-23
;  NUMBER OF SEQ ID NOS: 21907
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18790
;   LENGTH: 363
;   TYPE: DNA
;   ORGANISM: Mouse
US-09-724-671-18790


  Query Match             88.9%;  Score 16;  DB 7;  Length 363;
  Best Local Similarity   100.0%;  Pred. No. 5.6e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 gcagcagcggcagaag 17
          ||||||||||||||||
Db     89 GCAGCAGCGGCAGAAG 74


Search completed: September 21, 2001, 02:39:19
Job time: 15045 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: September 21, 2001, 00:27:43 ; Search time 3479.37 Seconds
(without alignments)
4338.867 Million cell updates/sec

Title: US-09-138-735-1_COPY_1232_2207
Perfect score: 976
Sequence: 1 caggtacagcgtaacggctt.....tgtgaacgaggccattatgt 976

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*
44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vi1:*
59: gb_vi2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	976	100.0	3402	9	AR047920	AR047920 Sequence
	2	972.8	99.7	3402	9	A48910	A48910 Sequence 1
	3	972.8	99.7	3402	96	TCU24190	U24190 Trypanosoma
	4	59.6	6.1	40309	6	LMFL1923	AL356246 Leishmani
	5	59.6	6.1	110000	84	LMFLCHR31_04	Continuation (5 of
	6	47.8	4.9	2402	3	MLURHO	L27277 Micrococcus
	7	46.6	4.8	175105	61	AC010531	AC010531 Homo sapi
c	8	46.6	4.8	179034	61	AC009109	AC009109 Homo sapi

	9	45	4.6	47852	3	MTV023	AL022022 Mycobacte
c	10	43.8	4.5	53662	3	MTV016	AL021841 Mycobacte
c	11	43.2	4.4	2028	15	SSI132828	AJ132828 Spermatoz
c	12	43.2	4.4	158390	60	AC008532	AC008532 Homo sapi
c	13	42.8	4.4	767	6	EIMMAX	M30933 E.tenella a
c	14	42.6	4.4	140812	13	AP001081	AP001081 Oryza sat
	15	41.6	4.3	5880	59	HSE4	D14486 Equine herp
c	16	41.2	4.2	2479	95	RNU87960	U87960 Rattus norv
	17	41.2	4.2	220374	60	AC008735	AC008735 Homo sapi
c	18	40.4	4.1	477	10	E34519	E34519 SCA7 gene a
c	19	40.4	4.1	477	88	AF020275	AF020275 Homo sapi
c	20	40.4	4.1	3932	93	HSM801297	AL133030 Homo sapi
c	21	40.4	4.1	3941	95	RNLAR2	X83546 R.norvegicu
c	22	40.4	4.1	5477	85	AB051453	AB051453 Homo sapi
	23	40.4	4.1	41230	3	MTCY253	Z81368 Mycobacteri
	24	40.4	4.1	69350	3	MTV004	AL009198 Mycobacte
c	25	40.4	4.1	149618	91	AP000556	AP000556 Homo sapi
	26	40.4	4.1	150036	91	AP000557	AP000557 Homo sapi
c	27	40.4	4.1	157086	91	AP000552	AP000552 Homo sapi
c	28	40.4	4.1	166447	87	AC018751	AC018751 Homo sapi
	29	40.4	4.1	169237	87	AC009516	AC009516 Homo sapi
	30	40.4	4.1	171350	72	AC058816	AC058816 Homo sapi
	31	40.4	4.1	176051	88	AC023490	AC023490 Homo sapi
c	32	40.4	4.1	212656	86	AC007957	AC007957 Homo sapi
	33	40.2	4.1	701	13	AF310215	AF310215 Sorghum b
c	34	40.2	4.1	67200	3	MTV017	AL021897 Mycobacte
	35	40.2	4.1	138273	83	AP003297	AP003297 Oryza sat
c	36	40.2	4.1	172307	71	AC044842	AC044842 Homo sapi
	37	40	4.1	624	5	AF139019	AF139019 Cepaea ne
	38	40	4.1	10732	10	E32986	E32986 Gene encodi
c	39	40	4.1	122395	78	AC090433	AC090433 Chlamydom
c	40	40	4.1	130569	78	AC090650	AC090650 Arabidops
	41	39.6	4.1	618	14	CRCGCR1	X17207 Chlamydomon
c	42	39.4	4.0	947	53	CNS0794Y	AL434840 T3 end of
	43	39.4	4.0	1442	88	AF087653	AF087653 Homo sapi
	44	39.4	4.0	29924	3	SC5E9	AL446003 Streptomy
c	45	39.2	4.0	818	8	CHKPROTAMI	L38713 Gallus gall

ALIGNMENTS

```
RESULT   1
AR047920
LOCUS       AR047920      3402 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 1 from patent US 5820864.
ACCESSION   AR047920
VERSION     AR047920.1  GI:5970263
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 3402)
  AUTHORS   Paranhos-Baccala,G., Lesenechal,M. and Jolivet,M.
  TITLE     Trypanosoma cruzi antigen, gene encoding therefor and methods of
            detecting and treating chagas disease
  JOURNAL   Patent: US 5820864-A 1 13-OCT-1998;
FEATURES             Location/Qualifiers
     source          1. .3402
                     /organism="unknown"
BASE COUNT      888 a    821 c    956 g    737 t
ORIGIN
```

Query Match 100.0%; Score 976; DB 9; Length 3402;
Best Local Similarity 100.0%; Pred. No. 1.6e-203;
Matches 976; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 caggtacagcgtaacggcttttgcttcaatcgtacagccgacggtagctgcgtcctggct 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1232 CAGGTACAGCGTAACGGCTTTTGCTTCAATCGTACAGCCGACGGTAGCTGCGTCCTGGCT 1291

Qy      61 gacatgtcgattcgattgacgatcttccatctccggtcctcccgcagggaagaacagcag 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1292 GACATGTCGATTCGATTGACGATCTTCCATCTCCGGTCCTCCCGCAGGGAAGAACAGCAG 1351

Qy     121 ccaggccaaaaaacatcggtagtggcgacggcgaaaccggggtgtgtgtcctcgggcact 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1352 CCAGGCCAAAAAACATCGGTAGTGGCGACGGCGAAACCGGGGTGTGTGTCCTCGGGCACT 1411

Qy     181 gacgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatcccctgcatca 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1412 GACGCGGCGAGTAGCAGTCATACCAATACGACTTCTGCCGCTGCTGCATCCCCTGCATCA 1471

Qy     241 ccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggctgag 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1472 CCCCCTGTTTCAGCGCCAGCCAAGGCAGCCGCGCCTCCTGCCGCGGCGCGATCGGCTGAG 1531

Qy     301 ccgcacgtggggagcaagatcattgctaatctagtgaatcagctggggattaatgtcacc 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1532 CCGCACGTGGGGAGCAAGATCATTGCTAATCTAGTGAATCAGCTGGGGATTAATGTCACC 1591

Qy     361 caaaggagcgtcgtcagcactggagcgccggccacgacgaggtctacggcggtgacgtcc 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1592 CAAAGGAGCGTCGTCAGCACTGGAGCGCCGGCCACGACGAGGTCTACGGCGGTGACGTCC 1651

Qy     421 acgactaccgccccgcagcgaacaagtccatacgggcacaatggccgacctgtgacggct 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1652 ACGACTACCGCCCCGCAGCGAACAAGTCCATACGGGCACAATGGCCGACCTGTGACGGCT 1711

Qy     481 ggattggtggcagctaatagtggtgccagcgcggcctcgtctcccacagccgcggcgaaa 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1712 GGATTGGTGGCAGCTAATAGTGGTGCCAGCGCGGCCTCGTCTCCCACAGCCGCGGCGAAA 1771

Qy     541 ccaacaggagaagaaaaggcctccgcggcatgtgaaacgagctccgtggcgataaatgcg 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1772 CCAACAGGAGAAGAAAAGGCCTCCGCGGCATGTGAAACGAGCTCCGTGGCGATAAATGCG 1831

Qy     601 acacgcccggcgcttcacaacgcctctctcccgcaggcgccaacggatggcgttttggcg 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1832 ACACGCCCGGCGCTTCACAACGCCTCTCTCCCGCAGGCGCCAACGGATGGCGTTTTGGCG 1891

Qy     661 gcagcagtataccagtcggagggcgaggttcatcagtcgctggagcggctggagtccgtc 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1892 GCAGCAGTATACCAGTCGGAGGGCGAGGTTCATCAGTCGCTGGAGCGGCTGGAGTCCGTC 1951

Qy     721 ataaccaacacgtctcgggttctgaagttgctccctgacaccattcgaagagaccatgaa 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1952 ATAACCAACACGTCTCGGGTTCTGAAGTTGCTCCCTGACACCATTCGAAGAGACCATGAA 2011

Qy     781 caacttctgaatctgggtttagaggcacagatgacagagctgcagcagagccgtccaaca 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2012 CAACTTCTGAATCTGGGTTTAGAGGCACAGATGACAGAGCTGCAGCAGAGCCGTCCAACA 2071

Qy     841 ccgcaaacacagccgagagacacaagctccgcgaaatcatccgtgtttgagacgtacacc 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2072 CCGCAAACACAGCCGAGAGACACAAGCTCCGCGAAATCATCCGTGTTTGAGACGTACACC 2131

Qy     901 cttgttctcattgcggattccctctctcgcaacatcacgaaggggtgaagcgtggtgtg 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2132 CTTGTTCTCATTGCGGATTCCCTCTCTCGCAACATCACGAAGGGGGTGAAGCGTGGTGTG 2191

Qy     961 aacgaggccattatgt 976
           ||||||||||||||||
Db    2192 AACGAGGCCATTATGT 2207
```

```
RESULT   2
A48910
LOCUS       A48910       3402 bp    DNA             PAT       07-MAR-1997
DEFINITION  Sequence 1 from Patent WO9605312.
ACCESSION   A48910
VERSION     A48910.1  GI:2302570
KEYWORDS    .
SOURCE      Trypanosoma cruzi.
  ORGANISM  Trypanosoma cruzi
```

```
           Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
           Trypanosoma; Schizotrypanum.
REFERENCE  1  (bases 1 to 3402)
  AUTHORS  Paranhos-Baccala,G., Lesenechal,M. and Jolivet,M.
  TITLE    NOVEL TRYPANOSOMA CRUZI ANTIGEN, AND GENE CODING THEREFOR; THEIR
           APPLICATION TO THE DETECTION OF CHAGAS' DISEASE
  JOURNAL  Patent: WO 9605312-A 1 22-FEB-1996;
           BIO MERIEUX (FR)
COMMENT    Other publication CA 2173957 960222
           Other publication AU 3169195 960307
           Other publication FR 2723589 960216.
FEATURES             Location/Qualifiers
     source          1. .3402
                     /organism="Trypanosoma cruzi"
                     /strain="G"
                     /db_xref="taxon:5693"
                     /dev_stage="EPIMASTIGOTE"
BASE COUNT      889 a    818 c    958 g    737 t
ORIGIN
```

Query Match 99.7%; Score 972.8; DB 9; Length 3402;
Best Local Similarity 99.8%; Pred. No. 7.8e-203;
Matches 974; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy       1 caggtacagcgtaacggctttgcttcaatcgtacagccgacggtagctgcgtcctggct 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1232 CAGGTACAGCGTAACGGCTTTTGCTTCAATCGTACAGCCGACGGTAGCTGCGTCCTGGCT 1291

Qy      61 gacatgtcgattcgattgacgatcttccatctccggtcctcccgcagggaagaacagcag 120
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1292 GACATGTCGAATCGATTGACGATCTTCCATCTCCGGTCCTCCCGCAGGGAAGAACAGCAG 1351

Qy     121 ccaggccaaaaaacatcggtagtggcgacggcgaaaccggggtgtgtgtcctcgggcact 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1352 CCAGGCCAAAAAACATCGGTAGTGGCGACGGCGAAACCGGGGTGTGTGTCCTCGGGCACT 1411

Qy     181 gacgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatcccctgcatca 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1412 GACGCGGCGAGTAGCAGTCATACCAATACGACTTCTGCCGCTGCTGCATCCCCTGCATCA 1471

Qy     241 ccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggctgag 300
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    1472 CCCCCTGTTTCAGCGCCAGCGAAGGCAGCCGCGCCTCCTGCCGCGGCGCGATCGGCTGAG 1531

Qy     301 ccgcacgtggggagcaagatcattgctaatctagtgaatcagctggggattaatgtcacc 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1532 CCGCACGTGGGGAGCAAGATCATTGCTAATCTAGTGAATCAGCTGGGGATTAATGTCACC 1591

Qy     361 caaaggagcgtcgtcagcactggagcgccggccacgacgaggtctacggcggtgacgtcc 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1592 CAAAGGAGCGTCGTCAGCACTGGAGCGCCGGCCACGACGAGGTCTACGGCGGTGACGTCC 1651

Qy     421 acgactaccgccccgcagcgaacaagtccatacgggcacaatggccgacctgtgacggct 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1652 ACGACTACCGCCCCGCAGCGAACAAGTCCATACGGGCACAATGGCCGACCTGTGACGGCT 1711

Qy     481 ggattggtggcagctaatagtggtgccagcgcggcctcgtctcccacagccgcggcgaaa 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1712 GGATTGGTGGCAGCTAATAGTGGTGCCAGCGCGGCCTCGTCTCCCACAGCCGCGGCGAAA 1771

Qy     541 ccaacaggagaagaaaaggcctccgcggcatgtgaaacgagctccgtggcgataaatgcg 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1772 CCAACAGGAGAAGAAAAGGCCTCCGCGGCATGTGAAACGAGCTCCGTGGCGATAAATGCG 1831

Qy     601 acacgcccggcgcttcacaacgcctctctcccgcaggcgccaacggatggcgttttggcg 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1832 ACACGCCCGGCGCTTCACAACGCCTCTCTCCCGCAGGCGCCAACGGATGGCGTTTTGGCG 1891

Qy     661 gcagcagtataccagtcggagggcgaggttcatcagtcgctggagcggctggagtccgtc 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1892 GCAGCAGTATACCAGTCGGAGGGCGAGGTTCATCAGTCGCTGGAGCGGCTGGAGTCCGTC 1951

Qy     721 ataaccaacacgtctcgggttctgaagttgctccctgacaccattcgaagagaccatgaa 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1952 ATAACCAACACGTCTCGGGTTCTGAAGTTGCTCCCTGACACCATTCGAAGAGACCATGAA 2011

Qy     781 caacttctgaatctgggtttagaggcacagatgacagagctgcagcagagccgtccaaca 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2012 CAACTTCTGAATCTGGGTTTAGAGGCACAGATGACAGAGCTGCAGCAGAGCCGTCCAACA 2071

Qy     841 ccgcaaacacagccgagagacacaagctccgcgaaatcatccgtgtttgagacgtacacc 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2072 CCGCAAACACAGCCGAGAGACACAAGCTCCGCGAAATCATCCGTGTTTGAGACGTACACC 2131

Qy     901 cttgttctcattgcggattccctctctcgcaacatcacgaaggggtgaagcgtggtgtg 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2132 CTTGTTCTCATTGCGGATTCCCTCTCTCGCAACATCACGAAGGGGGTGAAGCGTGGTGTG 2191

Qy     961 aacgaggccattatgt 976
          ||||||||||||||||
Db    2192 AACGAGGCCATTATGT 2207
```

```
RESULT   3
TCU24190
LOCUS       TCU24190     3402 bp    mRNA            INV       04-AUG-1997
DEFINITION  Trypanosoma cruzi Tc40 antigen (Tc40) mRNA, complete cds.
ACCESSION   U24190
VERSION     U24190.1  GI:790645
KEYWORDS    .
SOURCE      Trypanosoma cruzi.
  ORGANISM  Trypanosoma cruzi
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma; Schizotrypanum.
REFERENCE   1  (bases 1 to 3402)
  AUTHORS   Lesenechal,M., Duret,L., Cano,M.I., Mortara,R.A., Jolivet,M.,
            Camargo,M.E., da Silveira,J.F. and Paranhos-Baccala,G.
  TITLE     Cloning and characterization of a gene encoding a novel
            immunodominant antigen of Trypanosoma cruzi
  JOURNAL   Mol. Biochem. Parasitol. 87 (2), 193-204 (1997)
  MEDLINE   97391123
REFERENCE   2  (bases 1 to 3402)
  AUTHORS   Lesenechal,M., Franco Da Silveira,J., Mortara,R.A., Duret,L.,
            Camargo,M.E., Jolivet,M. and Paranhos-Baccala,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-APR-1995) Mylene Lesenechal, BioMerieux, 22 rue Saint
            Jean de Dieu, Lyon 69007, France
FEATURES             Location/Qualifiers
     source          1. .3402
                     /organism="Trypanosoma cruzi"
                     /strain="G"
                     /db_xref="taxon:5693"
                     /dev_stage="epimastigote"
     misc_RNA        1. .35
                     /gene="Tc40"
                     /note="spliced leader"
     gene            1. .3402
                     /gene="Tc40"
     CDS             266. .3013
                     /gene="Tc40"
                     /codon_start=1
                     /product="Tc40 antigen"
                     /protein_id="AAC47657.1"
                     /db_xref="GI:790646"
                     /translation="MTVTVDLFNHAKPSNNEGRVWSVDAATFNEVPEAQRVLADSQFY
                     LAYTMKRRHVLRVVKRSNLLKGTVRAHSKPIHAVKFVNYRSNVAASAGKGEFFVWVVT
                     DETEASNGKPDLAARLTVKVYFKLQDPVTIPCFSFFINAESQRPDLLVLYETQAAILD
                     SSSLIERFDVESLEATLQRNCTTLRTLTQPVSENSLCSVGSGGWFTFTTEPTMVAACT
                     LRNRSTPSWACCEGEPVKALHLLDATVEENVSVLVAASTKGVYQWLLTGVAEPNLLRK
                     FVIDGSIVAMESSRETFAVFDDRKQLALVNMHSPHNFTCTHYMMPCQVQRNGFCFNRT
                     ADGSCVLADMSNRLTIFHLRSSRREEQQPGQKTSVVATAKPGCVSSGTDAASSSHTNT
                     TSAAAASPASPPVSAPAKAAAPPAAARSAEPHVGSKIIANLVNQLGINVTQRSVVSTG
                     APATTRSTAVTSTTTAPQRTSPYGHNGRPVTAGLVAANSGASAASSPTAAAKPTGEEK
                     ASAACETSSVAINATRPALHNASLPQAPTDGVLAAAVYQSEGEVHQSLERLESVITNT
                     SRVLKLLPDTIRRDHEQLLNLGLEAQMTELQQSRPTPQTQPRDTSSAKSSVFETYTLV
```

```
                     LIADSLSRNITKGVKRGVNEAIMLHLDHEVRHAIGNRLRQTQKNIIKSRLDEALKEST
                     TQFTAQLTQTVENLVKRELAEVLGSINGSLTSLVKENASLQKELNSIMSSGVLDEMRR
                     MREELCTLRESVAKRKATMPDSSLHATSSFQGRRSAPETILATALSMVREQQYRQGLE
                     YMLMAQQPSLLLRFLSILTRENENAYSELIENVETPNDVWCSVLLQLIEAAATEAEKE
                     VVVGVAIDILSERDQIAQNGALGSKLTTAMRAFERQARSETTSRSFLQCLKNLEKLLQ
                     S"
BASE COUNT      889 a    818 c    958 g    737 t
ORIGIN


  Query Match             99.7%;  Score 972.8;  DB 96;  Length 3402;
  Best Local Similarity   99.8%;  Pred. No. 7.8e-203;
  Matches  974;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 caggtacagcgtaacggcttttgcttcaatcgtacagccgacggtagctgcgtcctggct 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1232 CAGGTACAGCGTAACGGCTTTTGCTTCAATCGTACAGCCGACGGTAGCTGCGTCCTGGCT 1291

Qy     61 gacatgtcgattcgattgacgatcttccatctccggtcctcccgcagggaagaacagcag 120
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1292 GACATGTCGAATCGATTGACGATCTTCCATCTCCGGTCCTCCCGCAGGGAAGAACAGCAG 1351

Qy    121 ccaggccaaaaaacatcggtagtggcgacggcgaaaccggggtgtgtgtcctcgggcact 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1352 CCAGGCCAAAAAACATCGGTAGTGGCGACGGCGAAACCGGGGTGTGTGTCCTCGGGCACT 1411

Qy    181 gacgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatcccctgcatca 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1412 GACGCGGCGAGTAGCAGTCATACCAATACGACTTCTGCCGCTGCTGCATCCCCTGCATCA 1471

Qy    241 ccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggctgag 300
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   1472 CCCCCTGTTTCAGCGCCAGCGAAGGCAGCCGCGCCTCCTGCCGCGGCGCGATCGGCTGAG 1531

Qy    301 ccgcacgtggggagcaagatcattgctaatctagtgaatcagctggggattaatgtcacc 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1532 CCGCACGTGGGGAGCAAGATCATTGCTAATCTAGTGAATCAGCTGGGGATTAATGTCACC 1591

Qy    361 caaaggagcgtcgtcagcactggagcgccggccacgacgaggtctacggcggtgacgtcc 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1592 CAAAGGAGCGTCGTCAGCACTGGAGCGCCGGCCACGACGAGGTCTACGGCGGTGACGTCC 1651

Qy    421 acgactaccgccccgcagcgaacaagtccatacgggcacaatggccgacctgtgacggct 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1652 ACGACTACCGCCCCGCAGCGAACAAGTCCATACGGGCACAATGGCCGACCTGTGACGGCT 1711

Qy    481 ggattggtggcagctaatagtggtgccagcgcggcctcgtctcccacagccgcggcgaaa 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1712 GGATTGGTGGCAGCTAATAGTGGTGCCAGCGCGGCCTCGTCTCCCACAGCCGCGGCGAAA 1771

Qy    541 ccaacaggagaagaaaaggcctccgcggcatgtgaaacgagctccgtggcgataaatgcg 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1772 CCAACAGGAGAAGAAAAGGCCTCCGCGGCATGTGAAACGAGCTCCGTGGCGATAAATGCG 1831

Qy    601 acacgcccggcgcttcacaacgcctctctcccgcaggcgccaacggatggcgttttggcg 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1832 ACACGCCCGGCGCTTCACAACGCCTCTCTCCCGCAGGCGCCAACGGATGGCGTTTTGGCG 1891

Qy    661 gcagcagtataccagtcggagggcgaggttcatcagtcgctggagcggctggagtccgtc 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1892 GCAGCAGTATACCAGTCGGAGGGCGAGGTTCATCAGTCGCTGGAGCGGCTGGAGTCCGTC 1951

Qy    721 ataaccaacacgtctcgggttctgaagttgctccctgacaccattcgaagagaccatgaa 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1952 ATAACCAACACGTCTCGGGTTCTGAAGTTGCTCCCTGACACCATTCGAAGAGACCATGAA 2011

Qy    781 caacttctgaatctgggtttagaggcacagatgacagagctgcagcagagccgtccaaca 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2012 CAACTTCTGAATCTGGGTTTAGAGGCACAGATGACAGAGCTGCAGCAGAGCCGTCCAACA 2071

Qy    841 ccgcaaacacagccgagagacacaagctccgcgaaatcatccgtgtttgagacgtacacc 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2072 CCGCAAACACAGCCGAGAGACACAAGCTCCGCGAAATCATCCGTGTTTGAGACGTACACC 2131

Qy    901 cttgttctcattgcggattccctctctcgcaacatcacgaaggggtgaagcgtggtgtg 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2132 CTTGTTCTCATTGCGGATTCCCTCTCTCGCAACATCACGAAGGGGGTGAAGCGTGGTGTG 2191

Qy    961 aacgaggccattatgt 976
          ||||||||||||||||
Db   2192 AACGAGGCCATTATGT 2207


RESULT   4
LMFL1923
LOCUS       LMFL1923    40309 bp    DNA             INV       16-MAY-2000
DEFINITION  Leishmania major Friedlin chromosome 14 cosmid L1923.
ACCESSION   AL356246
VERSION     AL356246.1  GI:7940262
KEYWORDS    possible DNA polymerase I-related protein; possible glutathione
            synthetase; possible ribosomal protein; possible TC40 antigen;
            protein phosphatase 2C-like protein.
SOURCE      Leishmania major.
  ORGANISM  Leishmania major
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.
REFERENCE   1  (bases 1 to 40309)
  AUTHORS   Ivens,A.C., Lewis,S.M., Bagherzadeh,A., Zhang,L., Chan,H.M. and
            Smith,D.F.
  TITLE     A physical map of the Leishmania major Friedlin genome
  JOURNAL   Genome Res. 8 (2), 135-145 (1998)
  MEDLINE   98146435
REFERENCE   2  (bases 1 to 40309)
  AUTHORS   Murphy,L., Harris,D., Ivens,A.C., Quail,M., Rajandream,M.A. and
            Barrell,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2000) European Leishmania major Friedlin genome
            sequencing project, Sanger Centre, The Wellcome Trust Genome
            Campus, Hinxton, Cambridge CB10 1SA, (E-mail: barrell@sanger.ac.uk)
COMMENT     see http://www.ebi.ac.uk/parasites/leish.html
            Notes:
            Details of leishmania sequencing at the Sanger Centre are available
            on the World Wide Web.
            see http://www.sanger.ac.uk/Projects/L_major/
            CDS are numbered using the following system eg L1923.01. L1923
            (cosmid name), .01 (first CDS)
            To make the cosmid library Leishmania major Friedlin DNA was
            partially digested with Sau3AI prior to cloning into BamHI site of
            the cosmid shuttle vector cLHYG (Ryan et al. 1993 Gene
            131:145-150). The sequence of the packaged vector was determined by
            Peter Myler and Ken Stuart at Seattle Biomedical Research
            Institute, and is available as accession number U59231.
            The more significant matches with motifs in the PROSITE database
            are also included but some of these may be fortuitous.
            The length in codons is given for each CDS.
            Usually the highest scoring match found by fasta -o is given for
            CDS which show significant similarity to other CDS in the database.

            Gene prediction is done using:
            (1)
            the FramePlot program of Bibb et al.,
            Gene 30:157-166(1984) as implemented
            at http://www.nih.go.jp/
            jun/cgi-bin/frameplot.pl. (2)
            codon preference based on the codon usage table for Leishmania at
            http://www.kazusa.or.jp/codon/
            (3)
            the Hexamer program which was written by Richard Durbin as an
            integral part of the ACEDB-based analysis tools for the C.elegans
            Genome Sequencing Project. The program calculates the
            log-likelihood score for a given DNA segment based on the frequency
            of 6-mers, normalised for the base-pair composition of the genome.
            The program was trained using a fasta file of confirmed Leismania
            major coding sequences (CDS), i.e. from ATG start codon to the stop
            codon.
            CAUTION:  We may not have predicted the correct initiation codon.
```

```
          Where possible we choose an initiation codon (atg) which is
          preceded by a stretch of pyrimidines or part of a Kozak sequence.
          If this cannot be identified we choose the most upstream initiation
          codon.
          Transmembrane domains were predicted as implemented at the TMHMM
          server: http://www.cbs.dtu.dk/services/TMHMM-1.0/
          IMPORTANT: This sequence MAY NOT be the entire insert of the
          sequenced clone.  It may be shorter because we only sequence
          overlapping sections once, or longer, because we arrange for a
          small overlap between neighbouring submissions.
          Cosmid L1923 is not overlapped by any cosmids sequenced to date.
FEATURES             Location/Qualifiers
     source          1. .40309
                     /organism="Leishmania major"
                     /strain="Friedlin"
                     /db_xref="taxon:5664"
                     /chromosome="14"
                     /clone="cosmid L1923"
     misc_feature    1. .209
                     /note="region of BLASTN similarity to: 233. .441 AQ852761
                     LMAJFV1_lm77b06.y1 Leishmania major FV1 random genomic
                     library Leishmania major genomic clone LMAJFV1_lm77b06 5'
                     similar to WP:F31C3.2 CE17731,contains 3. .462
                     V1-ch1_type_II.5 leishmania repetitive element, 99%
                     identity over 208 bases"
     gene            2. .1549
                     /gene="L1923.01"
     CDS             <2. .1549
                     /gene="L1923.01"
                     /note="L1923.01, len > 514 aa, unknown; some similarity to
                     P70431, enabled homolog (783 aa, Mus musculus, EMBL:
                     U72521, AAC52864); Fasta scores: E():0.63, 26.4% identity
                     in 182 aa"
                     /codon_start=1
                     /label=L1923.01
                     /product="hypothetical protein L1923.01"
                     /protein_id="CAB91830.1"
                     /db_xref="GI:7940263"
                     /translation="IYYLVHRGVIPYVCPTRDIPASLRRNVDADPQYAAMVDPAWTPE
                     ERAAMEAQAGELLLGFFYYYAFEFDWANHVVSLNRPGVTTKAALGWDVEDVAQASGSV
                     PHFGVTGWQYQHNLAGAEGQHGDAHGGTAPAASQTRPSVGHDGMLASNASMSARRSRA
                     TTRYSFCIEDPYEENLNLGRHMGVTKTLRVHTELYRGLLSLLKDDLQHCCVFAASVNS
                     SGSTGSNAATVSSAAEPATAAAKTSAEPTELPVRVLYKLMAVSTRELAMARRRHSAAV
                     TAGAEFPGVWLSDLEAAFLAQAPTEWKLATQVWNKHQLLHRLGLKLHAREYVLPRREV
                     GVRRLAAKAPPGVVLASTPEPTFNAEEVAAAAAEGERAPSSTEHAPVSSAEVTQMNQA
                     FLGALPVRRLPEDLLLAMTKGYSCLTPSWVAWTKPWAALSAWWTDRLHSPSTTPQSEG
                     PLASGASEQGGVIPSLPPGAPRHISAMPEKSAGAMHQTRTKLRRHVVAETASVSAGRR
                     ALRMLFR"
     misc_feature    complement(416. .870)
                     /note="region of BLASTN similarity to: 3. .457 AZ048344
                     LMAJFV1_lm77b06.x1 Leishmania major FV1 random genomic
                     library Leishmania major genomic clone LMAJFV1_lm77b06 3,
                     99% identity over 454 bases"
     repeat_region   1880. .1900
                     /note="(cca)7"
                     /label=(cca)7
     repeat_region   2047. .2058
                     /note="(aca)4"
                     /label=(aca)4
     repeat_region   2060. .2074
                     /note="poly-pyrimidine tract"
                     /label=ccccccccccccccc
     repeat_region   complement(2075. .2094)
                     /note="poly-pyrimidine tract"
                     /label=cttcttcccttttttcttcct
     repeat_region   2385. .2401
                     /note="poly-pyrimidine tract"
                     /label=ccccctttcccctccctc
     repeat_region   2749. .2763
                     /note="poly-pyrimidine tract"
                     /label=ttttctttttttttct
     gene            2880. .3386
                     /gene="L1923.02"
     CDS             2880. .3386
                     /gene="L1923.02"
                     /note="L1923.02, len = 167 aa, possibly ribosomal protein;
                     reasonable similarity to several, e.g. RL16_ACYKS, 50S
                     ribosomal protein L16 (136 aa, Acyrthosiphon kondoi
                     symbiotic bacterium, EMBL: D31786, BAA06582); Fasta
                     scores: E():1e-05, 31.7% identity in 142 aa"
                     /codon_start=1
                     /label=L1923.02
                     /product="possible ribosomal protein"
                     /protein_id="CAB91831.1"
                     /db_xref="GI:7940264"
                     /translation="MQSSALRPSSLLRAYHYVHPKTGRPIRQYRKYGDPRFINNEVAL
                     IKGLFALITTDFGMVTQSQLENARLAILRRMPRGSFALTMHTDYEEFPVVQKSPESRM
                     GAGKSNIHHFAYKFTTGVPLFEITALSSRRLNQAEAEGIFLAGRSFIPLQTAVVPRGR
                     VDEYHVFK"
     misc_feature    complement(3030. .3385)
                     /note="region of BLASTN similarity to: 166. .521 AQ946047
                     Sheared DNA-35J14.TR Sheared DNA Trypanosoma brucei
                     genomic clone Sheared DNA-35J1, 76% identity over 355
                     bases"
     misc_feature    3886. .4005
                     /note="region of BLASTN similarity to: 120529. .120648
                     AF008205 Leishmania major chromosome 1, complete sequence,
                     63% identity over 119 bases"
     repeat_region   4120. .4137
                     /note="poly-pyrimidine tract"
                     /label=ctttcctttcttccttct
     repeat_region   4252. .4266
                     /note="poly-pyrimidine tract"
                     /label=cctccctctcctccc
     misc_feature    4548. .4668
                     /note="region of BLASTN similarity to: 1. .121 AQ853285
                     LMAJFV1_lm82e05.y1 Leishmania major FV1 random genomic
                     library Leishmania major genomic clone LMAJFV1_lm82e05 5,
                     100% identity over 120 bases"
     repeat_region   4564. .4581
                     /note="poly-pyrimidine tract"
                     /label=ctccctttctctcttcct
     repeat_region   4688. .4707
                     /note="(gcccc)4"
                     /label=(gcccc)4
     repeat_region   4839. .4855
                     /note="poly-pyrimidine tract"
                     /label=tctctcccctccttccc
     repeat_region   4916. .4932
                     /note="poly-pyrimidine tract"
                     /label=ccccctttctcctccttt
     misc_feature    complement(5388. .5916)
                     /note="region of BLASTN similarity to: 1. .529 AQ901878
                     LMAJFV1_lm82e05.x1 Leishmania major FV1 random genomic
                     library Leishmania major genomic clone LMAJFV1_lm82e05 3,
                     99% identity over 528 bases"
     gene            5955. .8579
                     /gene="L1923.03"
     CDS             5955. .8579
                     /gene="L1923.03"
                     /note="L1923.03, len = 873 aa, unknown; contains PROSITE
                     PS00152 ATP synthase alpha and beta subunits signature"
                     /codon_start=1
                     /label=L1923.03
                     /product="hypothetical protein L1923.03"
                     /protein_id="CAB91832.1"
                     /db_xref="GI:7940265"
                     /translation="MLLVSASSFRGTQHGAHRHHSAHHQSSALYGAGTTTPDTSNLYQ
                     SLGSINLTVSANSNNPYVGVRMTPEATLLQPQSVAHGRSAEAGPRTAVLEINRHLSFY

  Query Match             6.1%;  Score 59.6;  DB 6;  Length 40309;
  Best Local Similarity   51.2%;  Pred. No. 0.0017;
  Matches  165;  Conservative    0;  Mismatches  154;  Indels    3;  Gaps    1;

Qy    510 cgcggcctcgtctcccacagccgcggcgaaaccaacaggagaagaaaaggcctccgcggc 569
          | | | |   ||| |  |  |  |||| ||  ||||             |   ||||
```

```
Db  12568 CCCTGTCGGCTCTGCGTCCCCGCCGGCCAACGCAACGTCGTCGACCGCTACGAACGCGAA 12627

Qy    570 atgtgaaacgagctccgtggcgataaatgcgacacgcccggcgcttcacaacgcctctct 629
Db  12628 CCCGATAACTGCCATGGCCGCGTACAACAGCGCTCGCAGCGCTCTCCAAAACGCCACCCT 12687

Qy    630 cccgcaggcgccaacggatggcgttttggcggcagcagtataccagtcggagggcgaggt 689
Db  12688 GCCACAGTCGC---TGGACGGGCTTGTGGCGAGCGCCGTGTACCAGGGCGACGAGGCCTT 12744

Qy    690 tcatcagtcgctggagcggctggagtccgtcataaccaacacgtctcgggttctgaagtt 749
Db  12745 GCGCCGCGCCATCGATGGCCTGCTGAATGTTGTGAAGAACGTTACGCAGATCCTGTTGCT 12804

Qy    750 gctccctgacaccattcgaagagaccatgaacaacttctgaatctgggtttagaggcaca 809
Db  12805 CACCCCCGATCAGCTCACGCGCGATCACCAGCAGCTCGTCTCACTGGCGCTGGAGGCGCA 12864

Qy    810 gatgacagagctgcagcagagc 831
Db  12865 GATGACGGAGCTGCAGCAGACC 12886


RESULT   5
LMFLCHR31_04
WPCOMMENT
  Sequence split into 11 fragments  LOCUS LMFLCHR31 Accession AL499621
     Fragment Name      Begin     End
     LMFLCHR31_00           1    110000
     LMFLCHR31_01      100001    210000
     LMFLCHR31_02      200001    310000
     LMFLCHR31_03      300001    410000
     LMFLCHR31_04      400001    510000
     LMFLCHR31_05      500001    610000
     LMFLCHR31_06      600001    710000
     LMFLCHR31_07      700001    810000
     LMFLCHR31_08      800001    910000
     LMFLCHR31_09      900001   1010000
     LMFLCHR31_10     1000001   1027083
Continuation (5 of 11) of LMFLCHR31 from base 400001 (AL499621 Leishmania major chromosc


  Query Match             6.1%;  Score 59.6;  DB 84;  Length 110000;
  Best Local Similarity   51.2%;  Pred. No. 0.0012;
  Matches  165;  Conservative    0;  Mismatches  154;  Indels    3;  Gaps    1;

Qy    510 cgcggcctcgtctcccacagccgcggcgaaaccaacaggagaagaaaaggcctccgcggc 569
Db  63000 CCCTGTCGGCTCTGCGTCCCCGCCGGCCAACGCAACGTCGTCGACCGCTACGAACGCGAA 63059

Qy    570 atgtgaaacgagctccgtggcgataaatgcgacacgcccggcgcttcacaacgcctctct 629
Db  63060 CCCGATAACTGCCATGGCCGCGTACAACAGCGCTCGCAGCGCTCTCCAAAACGCCACCCT 63119

Qy    630 cccgcaggcgccaacggatggcgttttggcggcagcagtataccagtcggagggcgaggt 689
Db  63120 GCCACAGTCGC---TGGACGGGCTTGTGGCGAGCGCCGTGTACCAGGGCGACGAGGCCTT 63176

Qy    690 tcatcagtcgctggagcggctggagtccgtcataaccaacacgtctcgggttctgaagtt 749
Db  63177 GCGCCGCGCCATCGATGGCCTGCTGAATGTTGTGAAGAACGTTACGCAGATCCTGTTGCT 63236

Qy    750 gctccctgacaccattcgaagagaccatgaacaacttctgaatctgggtttagaggcaca 809
Db  63237 CACCCCCGATCAGCTCACGCGCGATCACCAGCAGCTCGTCTCACTGGCGCTGGAGGCGCA 63296

Qy    810 gatgacagagctgcagcagagc 831
Db  63297 GATGACGGAGCTGCAGCAGACC 63318


RESULT   6
MLURHO
```

```
LOCUS       MLURHO       2402 bp    DNA             BCT       12-NOV-1996
DEFINITION  Micrococcus luteus Rho factor (rho) gene, complete cds.
ACCESSION   L27277
VERSION     L27277.1  GI:1666539
KEYWORDS    .
SOURCE      Micrococcus luteus.
  ORGANISM  Micrococcus luteus
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Micrococcineae; Micrococcaceae; Micrococcus.
REFERENCE   1  (bases 905 to 2402)
  AUTHORS   Opperman,T. and Richardson,J.P.
  TITLE     Phylogenetic analysis of sequences from diverse bacteria with
            homology to the Escherichia coli rho gene
  JOURNAL   J. Bacteriol. 176 (16), 5033-5043 (1994)
  MEDLINE   94327472
REFERENCE   2  (bases 1 to 2402)
  AUTHORS   Nowatzke,W.L. and Richardson,J.P.
  TITLE     Characterization of an unusual Rho factor from the high G + C
            gram-positive bacterium Micrococcus luteus
  JOURNAL   J. Biol. Chem. 271 (2), 742-747 (1996)
  MEDLINE   96132802
REFERENCE   3  (bases 905 to 2402)
  AUTHORS   Richardson,J.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-FEB-1994) John P. Richardson, Chemistry, Indiana
            University, Bloomington, IN 47405, USA
REFERENCE   4  (bases 1 to 2402)
  AUTHORS   Nowatzke,W.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-NOV-1996) William L. Nowatzke, Chemistry, Indiana
            University, Bloomington, IN 47405, USA
COMMENT     On Nov 12, 1996 this sequence version replaced gi:968908.
FEATURES             Location/Qualifiers
     source          1. .2402
                     /organism="Micrococcus luteus"
                     /strain="EM"
                     /db_xref="taxon:1270"
     gene            288. .2363
                     /gene="rho"
     CDS             288. .2363
                     /gene="rho"
                     /function="transcription termination factor"
                     /codon_start=1
                     /transl_table=11
                     /product="Rho factor"
                     /protein_id="AAB18671.1"
                     /db_xref="GI:1666540"
                     /translation="MTESTEQTTPTNGGGLASLKLAQLQALASQLGIAGGSRMRKADL
                     VTAISDHQRGGSVADRDAAERAAQAPAAPAAETAPAAASSEDAAPAAERPARRRSRRA
                     DADTSAPAAAQDGQPQAEAREAQTEQAPRETASDQDRSGGSEARDEGEDRPQSERRSR
                     GRRRAGDDDAQQGQDRRSDGAQGEDGADADRRGDREDRDDNGRENGRGRNGRNGRDRD
                     NGRDRENGRENSRDRENGRDGSREQRGDKSEDGGRGDGGRGDRSRRDDRDDEGGRNRR
                     NRRNRNERGRNRRGRGGPEVDETELTEDDVLQPVAGILDVLDNYAFVRTSGYLPGPND
                     VYVSLAMVKKYGLRKGDAVVGPIAPRDGEKQQHHGGGSNRQKFNALVKISSVNGQPAV
                     EHPQRVEFGKLVPLYPQERLRLETDPKLIGPRVIDLVSPIGKGQRGLIVSPPKAGKTM
                     ILQSIANAIKTNNPEVHLMMVLVDERPEEVTDMQRSVDGEVIASTFDRPADDHTTLAE
                     LAIERAKRLVEMGRDVVVLLDSMTRLGRAYNLAAPASGRILSGGVDSSALYPPKKFFG
                     AARNIENGGSLTILATALVETGSRMDEVIFEEFKGTGNMELRLSRHLAERRIFPAVDV
                     NASGTRREEALLSQEEVKIMWKLRRVLSGLEQQQALDLLTNKIKDTASNAEFLMLVSK
                     TTLGSKGDD"
BASE COUNT      386 a    947 c    818 g    251 t
ORIGIN


  Query Match             4.9%;  Score 47.8;  DB 3;  Length 2402;
  Best Local Similarity   50.2%;  Pred. No. 1.5;
  Matches  118;  Conservative    0;  Mismatches  117;  Indels    0;  Gaps    0;

Qy    367 agcgtcgtcagcactggagcgccggccacgacgaggtctacggcggtgacgtccacgact 426
Db    799 AGCAGGGTCAGGACCGTCGCTCCGACGGCGCCCAGGGCGAGGACGGCGCCGACGCGGACC 858

Qy    427 accgccccgcagcgaacaagtccatacgggcacaatggccgacctgtgacggctggattg 486
```

```
          ||||  ||    | |  |                |   || | |   |||| | |  |
Db    859 GCCGCGGCGACCGGGAGGACCGTGACGACAACGGGCGCGAGAACGGCCGCGGCCGCAACG 918

Qy    487 gtggcagctaatagtggtgccagcgcggcctcgtctcccacagccgcggcgaaaccaaca 546
          |  ||| |      |    || |   ||||| || |  | | | | || |||| | || |
Db    919 GCCGCAACGGGCGGGACCGCGACAACGGCCGCGACCGCGAGAACGGCCGCGAGAACAGCC 978

Qy    547 ggagaagaaaaggcctccgcggcatgtgaaacgagctccgtggcgataaatgcga 601
          ||    |  |   |  ||||| |   |    |||||  || ||||| || | |||
Db    979 GGGACCGTGAGAACGGCCGCGACGGCTCGCGCGAGCAGCGCGGCGACAAGTCCGA 1033


RESULT   7
AC010531
LOCUS       AC010531   175105 bp    DNA             HTG       31-AUG-2000
DEFINITION  Homo sapiens chromosome 16 clone RP11-178L8, WORKING DRAFT
            SEQUENCE, 2 ordered pieces.
ACCESSION   AC010531
VERSION     AC010531.5  GI:9954623
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 175105)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 175105)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Aug 31, 2000 this sequence version replaced gi:8576022.
            -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------.
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base-by-base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.
            ----------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   151748: contig of 151748 bp in length
            *  151749   151848: gap of unknown length
            *  151849   175105: contig of 23257 bp in length.
FEATURES            Location/Qualifiers
     source         1. .175105
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="16"
                    /clone="RP11-178L8"
BASE COUNT    47958 a  43943 c  43102 g  40002 t    100 others
ORIGIN


  Query Match             4.8%;  Score 46.6;  DB 61;  Length 175105;
  Best Local Similarity   49.4%;  Pred. No. 0.76;
  Matches  121;  Conservative    0;  Mismatches  124;  Indels    0;  Gaps    0;

Qy    538 aaaccaacaggagaagaaaaggcctccgcggcatgtgaaacgagctccgtggcgataaat 597
          ||| ||| |  | ||  |||    ||  || ||    || | |||   ||       || |
Db 125630 AAAACAAAACAAAAAACAAATCGGTCAACGTCACCAAAACCAAGCAGAGTCTAACAAACT 125689

Qy    598 gcgacacgcccggcgcttcacaacgcctctctcccgcaggcgccaacggatggcgttttg 657
          |  |||  || | |  |   ||| | | ||   | || | || ||   | |||||  || ||
Db 125690 GTCACAGCCCAGAGGAGGCACGAGGACTGAATGCCACGGGGGCTCCCAGATGGGATTCTG 125749

Qy    658 gcggcagcagtataccagtcggagggcgaggttcatcagtcgctggagcggctggagtcc 717
          |  | || |    |   |   ||   |   |   ||| |   |  | || |  |  ||
Db 125750 GGAGAAGAAAAGGATGTGAGTGAAAACTGCGGAAATCTGAACCAAGTGCAGGCGTGGTTA 125809

Qy    718 gtcataaccaacacgtctcgggttctgaagttgctccctgacaccattcgaagagaccat 777
          |||||||  |    | | || || | || |    |    |  |||| | ||||||  |
Db 125810 ATCATAACACATCATTATTGGTTTGTTAACTGTGACAAATATGCCATACTAAGAGAAGAC 125869

Qy    778 gaaca 782
          |  ||
Db 125870 GGTCA 125874


RESULT   8
AC009109/c
LOCUS       AC009109   179034 bp    DNA             HTG       31-AUG-2000
DEFINITION  Homo sapiens chromosome 16 clone RP11-467L24, WORKING DRAFT
            SEQUENCE, 14 ordered pieces.
ACCESSION   AC009109
VERSION     AC009109.8  GI:9954606
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 179034)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 179034)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Aug 31, 2000 this sequence version replaced gi:9256136.
            -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------.
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base-by-base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.
            ----------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1    23258: contig of 23258 bp in length
            *   23259    23358: gap of unknown length
            *   23359    31091: contig of 7733 bp in length
            *   31092    31191: gap of unknown length
```

```
            *    31192    66150: contig of 34959 bp in length
            *    66151    66250: gap of unknown length
            *    66251    71243: contig of 4993 bp in length
            *    71244    71343: gap of unknown length
            *    71344    84814: contig of 13471 bp in length
            *    84815    84914: gap of unknown length
            *    84915   100309: contig of 15395 bp in length
            *   100310   100409: gap of unknown length
            *   100410   107673: contig of 7264 bp in length
            *   107674   107773: gap of unknown length
            *   107774   109247: contig of 1474 bp in length
            *   109248   109347: gap of unknown length
            *   109348   125604: contig of 16257 bp in length
            *   125605   125704: gap of unknown length
            *   125705   144649: contig of 18945 bp in length
            *   144650   144749: gap of unknown length
            *   144750   174670: contig of 29921 bp in length
            *   174671   174770: gap of unknown length
            *   174771   176676: contig of 1906 bp in length
            *   176677   176776: gap of unknown length
            *   176777   177942: contig of 1166 bp in length
            *   177943   178042: gap of unknown length
            *   178043   179034: contig of 992 bp in length.
FEATURES             Location/Qualifiers
     source          1. .179034
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-467L24"
BASE COUNT    40736 a  43378 c  44428 g  49190 t   1302 others
ORIGIN

  Query Match             4.8%;  Score 46.6;  DB 61;  Length 179034;
  Best Local Similarity   49.4%;  Pred. No. 0.75;
  Matches  121;  Conservative    0;  Mismatches  124;  Indels    0;  Gaps    0;

Qy     538 aaaccaacaggagaagaaaaggcctccgcggcatgtgaaacgagctccgtggcgataaat 597
           ||| ||| |  | ||  |||    ||  || ||    || | |||   ||       || |
Db   48920 AAAACAAAACAAAAAACAAATCGGTCAACGTCACCAAAACCAAGCAGAGTCTAACAAACT 48861

Qy     598 gcgacacgcccggcgcttcacaacgcctctctcccgcaggcgccaacggatggcgttttg 657
           |  |||  || |  |   ||| | | ||   | || | || ||   | |||||  || ||
Db   48860 GTCACAGCCCAGAGGAGGCACGAGGACTGAATGCCACGGGGGCTCCCAGATGGGATTCTG 48801

Qy     658 gcggcagcagtataccagtcggagggcgaggttcatcagtcgctggagcggctggagtcc 717
           | | || |    |   |  ||   |   |   ||| |   |  | || |  |  ||
Db   48800 GGAGAAGAAAAGGATGTGAGTGAAAACTGCGGAAATCTGAACCAAGTGCAGGCGTGGTTA 48741

Qy     718 gtcataaccaacacgtctcgggttctgaagttgctccctgacaccattcgaagagaccat 777
           |||||||  |    | | || || | || |    |    |  |||| | ||||||  |
Db   48740 ATCATAACACATCATTATTGGTTTGTTAACTGTGACAAATATGCCATACTAAGAGAAGAC 48681

Qy     778 gaaca 782
           |  ||
Db   48680 GGTCA 48676


RESULT   9
MTV023
LOCUS       MTV023      47852 bp    DNA             BCT       17-JUN-1998
DEFINITION  Mycobacterium tuberculosis H37Rv complete genome; segment 148/162.
ACCESSION   AL022022 AL123456
VERSION     AL022022.1  GI:3261554
KEYWORDS    .
SOURCE      Mycobacterium tuberculosis.
  ORGANISM  Mycobacterium tuberculosis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE   1  (bases 1 to 47852)
  AUTHORS   Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
            Harris,D., Gordon,S.V., Eiglmeier,K., Gas,S., Barry III,C.E.,
            Tekaia,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
            Connor,R., Davies,R., Devlin,K., Feltwell,T., Gentles,S.,
            Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J.,
            Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
            Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S.,
            Squares,S., Sqares,R., Sulston,J.E., Taylor,K., Whitehead,S. and
            Barrell,B.G.
  TITLE     Deciphering the biology of Mycobacterium tuberculosis from the
            complete genome sequence
  JOURNAL   Nature 393 (6685), 537-544 (1998)
  MEDLINE   98295987
  REMARK    Erratum:[[published erratum appears in Nature 1998 Nov
            12;396(6707):190]]
REFERENCE   2  (bases 1 to 47852)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
            tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
            Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
            Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
            75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
COMMENT     On Jun 27, 1998 this sequence version replaced gi:2924430.
            Notes:
            Details of M. tuberculosis sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have
            been renumbered from the original cosmid submissions but the old
            gene designations are in brackets after the new gene numbers.
            Gene prediction was based on a Hidden Markov Model of TB genes
            implemented in TBparse (Krogh) supplemented with visual inspection
            of positional base preference in codons, especially where there is
            an increase in the observed/expected third position G + C.
            CAUTION:  In some cases we may not have predicted the correct
            initiation codon.  Where possible we choose an initiation codon
            (atg, gtg, or ttg) which is preceded by an upstream ribosome
            binding site sequence (optimally 5-13bp before the initiation
            codon).  If this cannot be identified we choose the most upstream
            initiation codon.
FEATURES             Location/Qualifiers
     source          1. .47852
                     /organism="Mycobacterium tuberculosis"
                     /strain="H37Rv"
                     /db_xref="taxon:1773"
     source          <1. .18
                     /organism="Mycobacterium tuberculosis"
                     /strain="H37Rv"
                     /db_xref="taxon:1773"
                     /clone="Y13E12"
     misc_feature    <1. .47852
                     /note="fragment designated v023. Does not represent a
                     physical clone"
     gene            complement(1. .1695)
                     /gene="Rv3494c"
     CDS             complement(1. .1695)
                     /gene="Rv3494c"
                     /note="Rv3494c, (MTV023.01c), len: 564. Unknown Pro-rich
                     protein similar to several Mycobacterium tuberculosis
                     proteins e.g. MTCI28.14 (515 aa), MTCY19H5.28c (516 aa)and
                     (MTV051.09). Has hydrophobic stretch, possibly signal
                     peptide at N-terminus. FASTA scores: Z97050|MTCI28_14
                     (515 aa)  opt: 1006 z-score: 582.8 E(): 5.9e-25;    34.3%
                     identity in 559 aa overlap; Z97182|MTCY19H5_28  (516 aa)
                     opt:  979 z-score:567.7 E(): 4.1e-24; 33.5% identity in
                     555 aa overlap. TBparse score is 0.897".
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein Rv3494c"
                     /protein_id="CAA17731.1"
                     /db_xref="GI:2924431"
                     /db_xref="SPTREMBL:O53539"
                     /translation="MIDRLAKIQLSIFAVITVITLSVMAIFYLRLPATFGIGTYGVSA
                     DFVAGGGLYKNANVTYRGVAVGRVESVGLNPNGVTAHMRLNSGTAIPSNVTATVRSVS
                     AIGEQYIDLVPPENPSSTKLRNGFRIQRQNTRIGQDVADLLRQAETLLGSLGDTRLRE
```

```
                LLHEAFIATNGAGPELARLIESARLLVDEANANYPQVSQLIDQAGPFLQAQIRAGGDI
                KSLADGLARFTWQLRAADPRLRDTLADAPDAIDEANTAFSGIRPSFPALAASLANLGR
                VGVIYHKSIEQLLVVFPALFAAIITSAGGVPQDEGAKLDFKIDLHDPPPCMTGFLPPP
                LVRSPADESVREIPRDMYCKTAQNDPSTVRGARNYPCQEFPGKRAPTVQLCRDPRGYV
                PVGTNPWRGPPIPYGTEVTDGRNILPPNKFPYIPPGADPDPGVPIVGPPPPGQVAGPG
                PAPHQPAQPAPPPNDNGPPPPFTSWMPPGYPPEPPQVPYPATIPPPPPPEGTGPPPGP
                APGPQPQASGPAYTIYDQLSGAFADPAGGTGIFAPGMTGASSAENWVDLMRDPRQL"
RBS             complement(1703. .1707)
                /note="possible RBS for Rv3494c"
gene            complement(1706. .2860)
                /gene="lprN"
CDS             complement(1706. .2860)
                /gene="lprN"
                /note="Rv3495c, (MTV023.02c), len: 384. lprN,  similar to
                Mycobacterium tuberculosis proteins MTCI28.13 (390 aa) and
                MTCY19H5.29 (402 aa) and (MTV051.08). Probably
                lipoprotein,contains possible signal sequence and
                appropriately positioned PS00013 Prokaryotic membrane
                lipoprotein lipid attachment site. FASTA scores:
                gp|Z97050|MTCI28_13 (390 aa)  opt:653 z-score: 762.1 E():
                0;    33.6% identity in 363 aa overlap; and
                Z97182|MTCY19H5_29 (402 aa)  opt: 572 z-score: 667.9 E():
                1.1e-29;    31.8% identity in 362 aa overlap. TBparse
                score is 0.897"
                /codon_start=1
                /transl_table=11
                /product="lprN"
                /protein_id="CAA17732.1"
                /db_xref="GI:2924432"
                /db_xref="SPTREMBL:O53540"
                /translation="MNRIWLRAIILTASSALLAGCQFGGLNSLPLPGTAGHGEGAYSV
                TVEMADVATLPQNSPVMVDDVTVGSVAGIVAVQRPDGSFYAAVKLDLDKNVLLPANAV
                AKVSQTSLLGSLHVELAPPTDRPPTGRLVDGSRITEANTDRFPTTEEVFSALGVVVNK
                GNVGALEEIIDETHQAVAGRQAQFVNLVPRLAELTAGLNRQVHDIIDALDGLNRVSAI
                LARDKDNLGRALDTLPDAVRVLNQNRDHIVDAFAALKRLTMVTSHVLAETKVDFGEDL
                KDLYSIVKALNDDRKDFVTSLQLLLTFPFPNFGIKQAVRGDYLNVFTTFDLTLRRIGE
                TFFTTAYFDPNMAHMDEILNPPDFLIGELANLSGQAADPFKIPPGTASGQ"
misc_feature    complement(2798. .2830)
                /gene="lprN"
                /note="PS00013 Prokaryotic membrane lipoprotein lipid
                attachment site"
gene            complement(2857. .4212)
                /gene="Rv3496c"
CDS             complement(2857. .4212)
                /gene="Rv3496c"
                /note="Rv3496c, (MTV023.03c), len: 451. Unknown but
                similar to Mycobacterium tuberculosis proteins MTCI28.12
                (530 aa), MTCY19H5.30c (508 aa) (MTV051.07). Hydrophobic
                region at N-terminus. FASTA scores: Z97050|MTCI28_12  (530
                aa) opt:838 z-score: 711.0 E(): 4.3e-32;    35.1% identity
                in 473 aaoverlap; and Z97182|MTCY19H5_30 (508 aa) opt:
                821 z-score:697.0 E(): 2.6e-31; 35.1% identity in 453 aa
                overlap.TBparse score is 0.891"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein Rv3496c"
                /protein_id="CAA17733.1"
                /db_xref="GI:2924433"
                /db_xref="SPTREMBL:O53541"
                /translation="MMGRVAMLTGSRGLRYATVIALVAALVGGVYVLSSTGNKRTIVG
                YFTSAVGLYPGDQVRVLGVPVGEIDMIEPRSSDVKITMSVSKDVKVPVDVQAVIMSPN
                LVAARFIQLTPVYTGGAVLPDNGRIDLDRTAVPVEWDEVKEGLTRLAADLSPAAGELQ
                GPLGAAINQAADTLDGNGDSLHNALRELAQVAGRLGDSRGDIFGTVKNLQVLVDALSE
                SDEQIVQFAGHVASVSQVLADSSANLDQTLGTLNQALSDIRGFLRENNSTLIETVNQL
                NDFAQTLSDQSENIEQVLHVAGPGITNFYNIYDPAQGTLNGLLSIPNFANPVQFICGG
                SFDTAAGPSAPDYYRRAEICRERLGPVLRRLTVNYPPIMFHPLNTITAYKGQIIYDTP
                ATEAKSETPVPELTWVPAGGGAPVGNPADLQSLLVPPAPGPAPAPPAPGAGPGEHGGG
                G"
RBS             complement(2867. .2871)
                /gene="Rv3496c"
                /note="possible RBS for Rv3495c"
gene            complement(4209. .5282)
                /gene="Rv3497c"
CDS             complement(4209. .5282)
                /gene="Rv3497c"
                /note="Rv3497c, (MTV023.04c), len: 357. Unknown but
                similar to Mycobacterium tuberculosis proteins MTCY19H5.31
                (481 aa), MTCI28.11, (515 aa)and MTV051.06. Hydrophobic
                region atN-terminus. FASTA scores: Z97182|MTCY19H5_31
                (481 aa) opt:611 z-score: 667.7 E(): 1.1e-29;    32.5%
                identity in 332aa overlap; and Z97050|MTCI28_11 (515 aa)
                opt:  587 z-score:641.2 E(): 3.3e-28;  30.1% identity in
                335 aa overlap. TBparse score is 0.889"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein Rv3497c"
                /protein_id="CAA17734.1"
                /db_xref="GI:2924434"
                /db_xref="SPTREMBL:O53542"
                /translation="MLNRKPSSKHERDPLRTGIFGLVLVICVVLIAFGYSGLPFWPQG
                KTYDAYFTDAGGITPGNSVYVSGLKVGAVSAVSLAGNSAKVTFSVDRSIVVGDQSLAA
                IRTDTILGERSIAVSPAGSGKSTTIPLSRTTTPYTLNGVLQDLGRNANDLNRPQFEQA
                LNVFTQALHDATPQVRGAVDGLTSLSRALNRRDEALQGLLAHAKSVTSVLSERAEQVN
                KLVEDGNQLFAALDARRAALSALISGIDDVAAQISGFVADNRKEFGPALSKLNLVLAN
                LNERRDYITEALKRLPTYATTLGEVVGSGPGFNVNVYSVLPGPLVATVFDLVFQPGKL
                PDSLADYLRGFIQERWIIRPKSP"
gene            complement(5272. .6324)
                /gene="Rv3498c"
CDS             complement(5272. .6324)
                /gene="Rv3498c"
                /note="Rv3498c, (MTV023.05c), len: 350. Unknown but
                similar to Mycobacterium tuberculosis proteins  MTCI28.10
                (346 aa), MTCY19H5.32c (346 aa) and MTV051.05. Hydrophobic
                region at N-terminus. FASTA scores: Z97050|MTCI28_10  (346
                aa)  opt: 755 z-score: 883.6 E():  0;    36.8% identity in
                340 aaoverlap; and Z97182|MTCY19H5_32 Mycobacterium
                tuberculosiscosmi (275 aa) opt:  699z-score: 820.1 E(): 0;
                47.0% identity in 249 aa overlap. TBparse score is 0.878"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein Rv3498c"
                /protein_id="CAA17735.1"
                /db_xref="GI:2924435"
                /db_xref="SPTREMBL:O53543"

  Query Match             4.6%;  Score 45;  DB 3;  Length 47852;
  Best Local Similarity   43.7%;  Pred. No. 2.5;
  Matches  198;  Conservative    0;  Mismatches  255;  Indels    0;  Gaps    0;

Qy    118 cagccaggccaaaaaacatcggtagtggcgacggcgaaaccggggtgtgtgtcctcgggc 177
          | | | ||| |     |  |||  | |  | |||| |    |   |  | | || ||| |
Db  32502 CCGGCGGGCAAGGCGGCGCCGGCGGCGATGGCGGCAACGGGGCCATCGGCGGCCACGGCC 32561

Qy    178 actgacgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatcccctgca 237
                | |||           ||     |  |  | ||  | || ||| |  | |||
Db  32562 CACTCACTGACGACCCCGGCGGCAACGGGGGCACCGGCGGCAACGGCGGCACCGGCGGCA 32621

Qy    238 tcacccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggct 297
          |  |  |     || || |||||  ||| || || |   | ||   ||||  | ||||
Db  32622 CCGGCGGCGCGGGCATCGGCAGCCTTGGCGGCGGCACTGGCGGCGATGGCGGCAACGGCG 32681

Qy    298 gagccgcacgtggggagcaagatcattgctaatctagtgaatcagctggggattaatgtc 357
          |   ||  ||  |     ||     || |     | |   | |  ||           ||
Db  32682 GCAACGGCGGTACCGGCGGCGAGGGCGGCGAGGTCGGCGGCGCCGGCGGCACCGGCGGTG 32741

Qy    358 acccaaaggagcgtcgtcagcactggagcgccggccacgacgaggtctacggcggtgacg 417
            |     |  |  |  |   | | | || ||||  || ||  |   ||| || | ||
Db  32742 CGGCCGGCAATGGCGGCGATGGCGGCACCGGCGGCACCGGCGGCGGGGACGGGGGCGCCG 32801

Qy    418 tccacgactaccgccccgcagcgaacaagtccatacgggcacaatggccgacctgtgacg 477
          |  |  |  | || |||  |  | |     | |  | |  |   ||     | |  ||
Db  32802 GCGGCACCGGCGGCACCGGCGGCACCGGCGGCCTCGGCGACCCCCGGGTCGGCGGATCCG 32861

Qy    478 gctggattggtggcagctaatagtggtgccagcgcggcctcgtctcccacagccgcggcg 537
          || |    || |||| |     |  ||    | |  |     |  || | |||||
Db  32862 GCGGCGACGGCGGCACCGGCGGCAGCGGCGGTGCGGCCGGCAATGGCGGCAACGGCGGCA 32921
```

```
Qy    538 aaaccaacaggagaagaaaaggcctccgcggca 570
          |  ||  |  | || | || |||  | ||||||
Db  32922 ACGCCGGCGCGGGAGGCAATGGCAACGGCGGCA 32954


RESULT  10
MTV016/c
LOCUS       MTV016      53662 bp    DNA             BCT       23-JUN-1999
DEFINITION  Mycobacterium tuberculosis H37Rv complete genome; segment 143/162.
ACCESSION   AL021841 AL123456
VERSION     AL021841.1  GI:3261517
KEYWORDS    .
SOURCE      Mycobacterium tuberculosis.
  ORGANISM  Mycobacterium tuberculosis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE   1  (bases 1 to 53662)
  AUTHORS   Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
            Harris,D., Gordon,S.V., Eiglmeier,K., Gas,S., Barry III,C.E.,
            Tekaia,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
            Connor,R., Davies,R., Devlin,K., Feltwell,T., Gentles,S.,
            Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J.,
            Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
            Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S.,
            Squares,S., Sqares,R., Sulston,J.E., Taylor,K., Whitehead,S. and
            Barrell,B.G.
  TITLE     Deciphering the biology of Mycobacterium tuberculosis from the
            complete genome sequence
  JOURNAL   Nature 393 (6685), 537-544 (1998)
  MEDLINE   98295987
  REMARK    Erratum:[[published erratum appears in Nature 1998 Nov
            12;396(6707):190]]
REFERENCE   2  (bases 1 to 53662)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
            tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
            Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
            Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
            75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
COMMENT     On Jun 27, 1998 this sequence version replaced gi:2894210.
            Notes:
            Details of M. tuberculosis sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have
            been renumbered from the original cosmid submissions but the old
            gene designations are in brackets after the new gene numbers.
            Gene prediction was based on a Hidden Markov Model of TB genes
            implemented in TBparse (Krogh) supplemented with visual inspection
            of positional base preference in codons, especially where there is
            an increase in the observed/expected third position G + C.
            CAUTION:  In some cases we may not have predicted the correct
            initiation codon.  Where possible we choose an initiation codon
            (atg, gtg, or ttg) which is preceded by an upstream ribosome
            binding site sequence (optimally 5-13bp before the initiation
            codon).  If this cannot be identified we choose the most upstream
            initiation codon.
FEATURES             Location/Qualifiers
     source          1. .53662
                     /organism="Mycobacterium tuberculosis"
                     /strain="H37Rv"
                     /db_xref="taxon:1773"
     misc_feature    <1. .50056
                     /note="fragment designated v016. Does not represent a
                     physical clone"
     RBS             complement(9. .13)
                     /note="possible RBS, aggag, for Rv2249c"
     gene            complement(15. .1496)
                     /gene="lpdA"
     CDS             complement(15. .1496)
                     /gene="lpdA"
                     /note="Rv3303c, (MTV016.02c), len: 493. Probable
                     dihydrolipoamide dehydrogenase similar to many e.g.
                     Q04829|DLDH_HALVO (474 aa) from HALOFERAX VOLCANII. FASTA
                     scores: sp|Q04829|DLDH_HALVO DIHYDROLIPOAMIDE
                     DEHYDROGENASE (E (474 aa) opt:675 z-score: 794.6 E(): 0;
                     29.3% identity in 471 aa overlap. TBparse score is 0.883"
                     /codon_start=1
                     /transl_table=11
                     /product="lpdA"
                     /protein_id="CAA17075.1"
                     /db_xref="GI:2894212"
                     /db_xref="SPTREMBL:O53355"
                     /translation="MVTRIVILGGGPAGYEAALVAATSHPETTQVTVIDCDGIGGAAV
                     LDDCVPSKTFIASTGLRTELRRAPHLGFHIDFDDAKISLPQIHARVKTLAAAQSADIT
                     AQLLSMGVQVIAGRGELIDSTPGLARHRIKATAADGSTSEHEADVVLVATGASPRILP
                     SAQPDGERILTWRQLYDLDALPDHLIVVGSGVTGAEFVDAYTELGVPVTVVASQDHVL
                     PYEDADAALVLEESFAERGVRLFKNARAASVTRTGAGVLVTMTDGRTVEGSHALMTIG
                     SVPNTSGLGLERVGIQLGRGNYLTVDRVSRTLATGIYAAGDCTGLLPLASVAAMQGRI
                     AMYHALGEGVSPIRLRTVAATVFTRPEIAAVGVPQSVIDAGSVAARTIMLPLRTNARA
                     KMSEMRHGFVKIFCRRSTGVVIGGVVVAPIASELILPIAVAVQNRITVNELAQTLAVY
                     PSLSGSITEAARRLMAHDDLDCTAAQDAAEQLALVPHHLPTSN"
     gene            1699. .2178
                     /gene="Rv3304"
     CDS             1699. .2178
                     /gene="Rv3304"
                     /note="Rv3304, (MTV016.03), len: 159. Unknown but very
                     similar to hypothetical protein L308_F3_97 (174 aa) from
                     M. leprae and some similarity to Q44546 (183 aa) from
                     Azotobacter vinelandii. ORF continues upstream with
                     possible start at2198. FASTA scores: sptr|Q49927|Q49927
                     L308_F3_97 (174 aa)opt:  974 z-score: 1017.7 E(): 0; 91.3%
                     identity in 149 aaoverlap and sptr|Q44546|Q44546
                     HYPOTHETICAL 20.3 KD PROTEIN. (183 aa) opt:  206 z-score:
                     227.5 E(): 3.5e-05; 29.0% identity in 145 aa overlap.
                     TBparse score is 0.922"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein Rv3304"
                     /protein_id="CAA17076.1"
                     /db_xref="GI:2894213"
                     /db_xref="SPTREMBL:O53356"
                     /translation="MPLYAAYGSNMHPEQMLERAPHSPMAGTGWLPGWRLTFGGEDIG
                     WEGALATVVEDPDSKVFVVLYDMTPADEKNLDRWEGSEFGIHQKIRCRVERISSDTTT
                     DPVLAWLYVLDAWEGGLPSARYLGVMADAAEIAGAPSDYVHDLRTRPARNIGPGTIA"
     gene            complement(2197. .3366)
                     /gene="amiA"
     CDS             complement(2197. .3366)
                     /gene="amiA"
                     /note="Rv3305c, (MTV016.04c), len: 389. amiA, Probable
                     N-acyl-L-amino acid amidohydrolase or peptidase similar to
                     many e.g. Bacillus subtilis BSUB0015_194 (416 aa) and
                     P54984|Y100_SYNY3 (393 aa). Contains PS00639 Eukaryotic
                     thiol (cysteine) proteases histidine active site. FASTA
                     scores: Z99118|BSUB0015_194  (416 aa) opt: 805 z-score:
                     953.3 E():  0; 37.1% identity in 364 aa overlap and
                     sp|P54984|Y100_SYNY3 HYPOTHETICAL 42.7 KD PROTEIN SLL0
                     (393 aa) opt:  802 z-score:950.2 E(): 0; 38.0% identity in
                     382 aa overlap. TBparse score is 0.905"
                     /codon_start=1
                     /transl_table=11
                     /product="amiA"
                     /protein_id="CAA17077.1"
                     /db_xref="GI:2894214"
                     /db_xref="SPTREMBL:O53357"
                     /translation="MSLADAAESWLAAHHDDLVGWRRHIHRYPELGRQEYATTQFVAE
                     RLADAGLNPKVLPGGTGLTCDFGPQHQPRIALRADMDALPMAERTGAPYASTMPNVAH
                     ACGHDAHTAILLGAALALASVPELPVGVRLIFQAAEELMPGGAIDAIAAGALAGVSRI
                     FALHCDPRLEVGKVAVRQGPITSAADSIEITLYSPGGHTSRPHLTADLVYGLGTLVTG
                     LPGVLSRRIDPRNSTVLVWGAVNAGMAANAIPQTGVLSGTVRTASRQTWVDLEELVRQ
                     AISALLLPLAIEHTLQYRRGVPPVVNEEISTRILAHAIEAIGPGVLADTRQSGGGEDF
                     SWYLEEVPGAMARLGVWSGDGLQLDLHQPTFDIDERALAIGLRVMVNIIEQAAAH"
     misc_feature    complement(2407. .2439)
                     /gene="amiA"
```

```
                /note="PS00639 Eukaryotic thiol (cysteine) proteases
                histidine active site"
     gene       complement(3363. .4547)
                /gene="amiB"
     CDS        complement(3363. .4547)
                /gene="amiB"
                /note="Rv3306c, (MTV016.05c), len: 394. amiB, Probable
                aminohydrolase similar to several belonging to peptidase
                family M40 (and to hypothetical proteins) e.g. AMHX_BACSU
                (383 aa). FASTA scores: sptr|P76052|P76052 (481 aa) opt:
                383 z-score: 455.1 E(): 7.3e-18; 30.5% identity in 328 aa
                overlap;sp|P44765|YDAJ_HAEIN HYPOTHETICAL PROTEIN HI0584
                (423 aa)opt:  297 z-score: 354.6 E(): 2.9e-12; 29.6%
                identity in 274aa overlap and sp|P54983|AMHX_BACSU
                AMIDOHYDROLASE AMHX (EC3.5.1.-)  (383 aa) opt:  294
                z-score: 351.7 E(): 4.2e-12; 26.6% identity in 379 aa
                overlap. TBparse score is 0.897"
                /codon_start=1
                /transl_table=11
                /product="amiB"
                /protein_id="CAA17078.1"
                /db_xref="GI:2894215"
                /db_xref="SPTREMBL:O53358"
                /translation="MPAASASDRVEELVRRRGGELVELSHAIHAEPELAFAEHRSCAK
                AQALVAERGFEITTAAGGLDTAFRADYGSGPLVVGVCAEYDALPGIGHACGHNIIAAS
                AVGTALALAEVADDLGLTVALLGTPAEESGGGKALMLQAGTFDDVAVAVMVHPGPTDI
                AGARSLALSEVTVRYRGKESHAAVAPHLGVNAADAVTVAQVAIGVLRQQLAPGQMVHG
                IVTDGGQAVNVIPGQARLQYAMRAVESDSLRELQTRMFACFAAGALAAGCEYEIDEAA
                PAYAELKPDPWLADVCREEMQRLGREPLLPALEAELPLGSTDMGNVTQVLPGIHPVIG
                LDAGAATVHQRAFTVASAGASADRAVVDGAIMLARTVVRLAQTPDERDRVLAAQQRRA
                AR"
     gene       4612. .5418
                /gene="deoD"
     CDS        4612. .5418
                /gene="deoD"
                /note="Rv3307, (MTV016.06), len: 268, deoD, probable
                purine nucleoside phosphorylase similar to several
                especially M.leprae DEOD_MYCLE (268 aa) and DEOD_BACST
                (274 aa). FASTA scores: sp|P46862|DEOD_MYCLE PROBABLE
                PURINE NUCLEOSIDE PHOSPH (268 aa) opt: 1373 z-score:
                1613.7 E(): 0; 82.1% identity in 262 aa overlap and
                sp|P77834|DEOD_BACST PURINE NUCLEOSIDE PHOSPHORYLASE (274
                aa) opt:  346 z-score: 413.5 E(): 1.5e-15;  42.0% identity
                in 245 aa overlap. TBparse score is0.878"
                /codon_start=1
                /transl_table=11
                /product="deoD"
                /protein_id="CAA17079.1"
                /db_xref="GI:2894216"
                /db_xref="SPTREMBL:O53359"
                /translation="MADPRPDPDELARRAAQVIADRTGIGEHDVAVVLGSGWLPAVAA
                LGSPTTVLPQAELPGFVPPTAAGHAGELLSVPIGAHRVLVLAGRIHAYEGHDLRYVVH
                PVRAARAAGAQIMVLTNAAGGLRADLQVGQPVLISDHLNLTARSPLVGGEFVDLTDAY
                SPRLRELARQSDPQLAEGVYAGLPGPHYETPAEIRMLQTLGADLVGMSTVHETIAARA
                AGAEVLGVSLVTNLAAGITGEPLSHAEVLAAGAASATRMGALLADVIARF"
     gene       5422. .7026
                /gene="pmmB"
     CDS        5422. .7026
                /gene="pmmB"
                /note="Rv3308, (MTV016.07), len: 534. Probable
                phosphomannomutase similar to several e.g. Mycobacterium
                leprae L308_F1_13 (534 aa) and sp|P18159|YHXB_BACSU (565

  Query Match             4.5%;  Score 43.8;  DB 3;  Length 53662;
  Best Local Similarity   43.3%;  Pred. No. 4.4;
  Matches  204;  Conservative    0;  Mismatches  267;  Indels    0;  Gaps    0;

Qy    142 gtggcgacggcgaaaccggggtgtgtgtcctcgggcactgacgcggcgagtagcagtcat 201
          | || |||||||    |||     |   | ||   ||  |  || |||
Db  48110 GCGGTGACGGCGTCGGCGGCAACAGCTCCGTCACCCAAGGCGGCAGCGGCGGTGGCGGCG 48051

Qy    202 accaatacgacttctgccgctgctgcatcccctgcatcacccctgtttcagcgccagcc 261
          |     || |  | || || || || |   | ||  ||    | | ||| ||| |  |
```

```
Db  48050 GCGCCGGCGGCGCCGGCGGCAGCGGCTTTTTCGGCGGCAAGGGCGGCTTCGGCGGCGACG 47991

Qy    262 aaggcagccgcgcctcctgccgcggcgcgatcggctgagccgcacgtggggagcaagatc 321
            ||     ||| | ||  | ||||||  | || | |  |||     || |     |
Db  47990 GCGGTCAGGGCGGCCCCAACGGCGGCGGTACCGTCGGCACCGTGGCCGGTGGCGGCGGCA 47931

Qy    322 attgctaatctagtgaatcagctggggattaatgtcacccaaaggagcgtcgtcagcact 381
          |  ||       | |      |  || ||           |    |   |  | |||  |
Db  47930 ACGGCGGTGTCGGCGGCCGGGGCGGCGACGGCGTCTTTGCCGGTGCCGGCGGCCAGGGCG 47871

Qy    382 ggagcgccggccacgacgaggtctacggcggtgacgtccacgactaccgccccgcagcga 441
          |   ||  || || | ||       ||||||   |  |  || | || ||  |   |
Db  47870 GCCTCGGTGGGCAGGGCGGCAATGGCGGCGGCTCCACCGGCGGCAACGGCGGCCTTGGCG 47811

Qy    442 acaagtccatacgggcacaatggccgacctgtgacggctggattggtggcagctaatagt 501
          |  |  |    | | |       |  |    | |||||    |||   | ||    ||
Db  47810 GCGCGGGCGGTGGCGGAGGCAACGCCCCGGACGGCGGCTTCGGTGGCAACGGCGGTAAGG 47751

Qy    502 ggtgccagcgcggcctcgtctcccacagccgcggcgaaaccaacaggagaagaaaaggcc 561
          |  ||||| ||||| | | |  |  ||  |   |||  |||  |    |  | | |||
Db  47750 GTGGCCAGGGCGGCATTGGCGGCGGCACTCAGAGCGCGACCGGCCTCGGCGGTGACGGCG 47691

Qy    562 tccgcggcatgtgaaacgagctccgtggcgataaatgcgacacgcccggcg 612
            ||||        ||    | | ||| | |   ||| || | ||||||
Db  47690 GTGACGGCGGTGACGGCGGCAACGGCGGCAACAGCGGCGCCAAGGCCGGCG 47640


RESULT  11
SSI132828/c
LOCUS       SSI132828    2028 bp    mRNA            PLN       05-APR-1999
DEFINITION  Spermatozopsis similis mRNA for p210 protein, partial.
ACCESSION   AJ132828
VERSION     AJ132828.1  GI:4584085
KEYWORDS    p210 protein.
SOURCE      Spermatozopsis similis.
  ORGANISM  Spermatozopsis similis
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;
            Spermatozopsis.
REFERENCE   1  (bases 1 to 2028)
  AUTHORS   Lechtreck,K.F., Teltenkoetter,A. and Grunow,A.
  TITLE     A 210-kDa-protein is located in a membrane-microtubule-linker at
            the distal end of mature and nascent basal bodies
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2028)
  AUTHORS   Lechtreck,K.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-FEB-1999) Lechtreck K.F., Botany, University of
            Cologne, Gyrhofstr. 15, D-50931 Cologne, GERMANY
FEATURES             Location/Qualifiers
     source          1. .2028
                     /organism="Spermatozopsis similis"
                     /db_xref="taxon:3192"
     CDS             <1. .1830
                     /function="putative microtubule-membrane-linker"
                     /codon_start=1
                     /product="p210 protein"
                     /protein_id="CAB40552.1"
                     /db_xref="GI:4584086"
                     /translation="APAHAPPPLPSPQVLAQQAARALSMAPPPPPSALQQAASQPAPP
                     AYAPPPQPASYAPQPAQQQAPAPPTYAPPPAQQQAPQAPAYAPPAPAQQQAPAPQAPQ
                     AYAPAPVAPAPQAPPPQQPPQQQQWAPAPAAPPGAALLQSAGALAGPAPSAGQVAAAQ
                     AVLSRQAAPAAAPQAPQAPQFAPPPQAPQFAPAPQQPQPQQQQPAAPQFAPPPQAPAP
                     QPQFAPAPQQQQALAPQFAPAPQQPQQQPAFAPAPQAPAPQAPAPQFAPQQPLAPQFA
                     PVPQAPAYAPAPQQPSYAPASPPAAGSPQPRPPAGSPVRPPAPTTESVLSADSEEGDG
                     SYSFEFDASEASAASPVRVKGGRSGSIGAAASGSAGAPGSRSAAGRYGDVQQSLEALS
                     PAPRAPAAPAAAAGPPVLRRLPADPAAWAAAGAGDDTLILSVESLTLSADALRDPRLG
                     EGAHVMVSHMLLEAFTAARDQTTAAAPARRGAIGFEYAVGYSLLARDGAPSEAGAALQ
                     RQVAVEGECIVPLVLLATTNPQVDLDVTSSGEFDMIGMAEVDLSDWFVNPKREDWVSR
                     DVRLSPHESAAAGYPSGVAGVVRVTLRARGALAALRRACGEGR"
BASE COUNT      233 a    808 c    797 g    189 t      1 others
ORIGIN
```

```
  Query Match             4.4%;  Score 43.2;  DB 15;  Length 2028;
  Best Local Similarity   43.8%;  Pred. No. 15;
  Matches  189;  Conservative    0;  Mismatches  243;  Indels    0;  Gaps    0;

Qy    137 cggtagtggcgacggcgaaaccggggtgtgtgtcctcgggcactgacgcggcgagtagca 196
          |||  |   || ||      |||| |   | | |  |||  |      || ||      |
Db    948 CGGGCGCACCGGCGAGCCGGCCGGCGGCCGCGGCTGCGGCGAGCCCGCCGCCGGCGGCGA 889

Qy    197 gtcataccaatacgacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgc 256
            |   |   ||||||  |||| || ||  |  | | |       | ||||   || ||
Db    888 CGCCGGCGCGTACGACGGCTGCTGCGGCGCCGGCGCGTAGGCGGGCGCCTGGGGCACCGG 829

Qy    257 cagccaaggcagccgcgcctcctgccgcggcgcgatcggctgagccgcacgtggggagca 316
          |    | || ||     |  |||| |||||| | | || | ||||      ||| |
Db    828 CGCGAACTGCGGCGCGAGCGGCTGCTGCGGCGCAAACTGCGGCGCCGGCGCCTGGGGGGC 769

Qy    317 agatcattgctaatctagtgaatcagctggggattaatgtcacccaaaggagcgtcgtca 376
          |     |||    |  | |     || ||    |  ||   |      |      |
Db    768 CGGCGCCTGCGGCGCCGGCGCGAACGCCGGCTGCTGTTGCGGCTGCTGCGGCGCCGGCGC 709

Qy    377 gcactggagcgccggccacgacgaggtctacggcggtgacgtccacgactaccgccccgc 436
          | ||||| ||||  ||  |  |     || |||||  | ||   ||     | ||  |||
Db    708 GAACTGGGGCGCAAGCGCCTGCTGCTGCTGCGGCGCCGGCGCGAACTGAGGCTGCGGCGC 649

Qy    437 agcgaacaagtccatacgggcacaatggccgacctgtgacggctggattggtggcagcta 496
          |   |     |    | ||  | ||     ||   | | ||||     | ||| ||
Db    648 GGGCGCCTGCGGCGGCGGCGCGAACTGCGGCGCCGCCGGCTGCTGCTGCTGCGGCTGCGG 589

Qy    497 atagtggtgccagcgcggcctcgtctcccacagccgcggcgaaaccaacaggagaagaaa 556
          |  ||  |      |||    |  |  |  | || |||| |   | ||| |  |
Db    588 CTGCTGCGGGGCCGGCGCAAACTGCGGCGCCTGCGGCGGTGGCGCGAACTGCGGCGCCTG 529

Qy    557 aggcctccgcgg 568
          |||  | ||||
Db    528 CGGCGCCTGCGG 517


RESULT  12
AC008532/c
LOCUS       AC008532   158390 bp    DNA             HTG       18-JUL-2000
DEFINITION  Homo sapiens chromosome 19 clone CTC-483I11, WORKING DRAFT
            SEQUENCE, 13 ordered pieces.
ACCESSION   AC008532
VERSION     AC008532.5  GI:9255996
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 158390)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 158390)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jul 18, 2000 this sequence version replaced gi:7689764.
            -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 370855, BC335879
            Center clone name: CIT-HSPC_483I11
            -------
            Summary Statistics
            Consensus quality: 145411 bases at least Q40
            Consensus quality: 154484 bases at least Q30
            Consensus quality: 156100 bases at least Q20
            Estimated insert size: 156970; agarose-fp estimation
            Estimated insert size: 157840; sum-of-contigs estimation
            Quality coverage:  5.76 in Q20 bases; agarose-fp estimation
            Quality coverage:  5.73 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1    10270: contig of 10270 bp in length
            *    10271    10370: gap of unknown length
            *    10371    17547: contig of 7177 bp in length
            *    17548    17647: gap of unknown length
            *    17648    62828: contig of 45181 bp in length
            *    62829    62928: gap of unknown length
            *    62929    74684: contig of 11756 bp in length
            *    74685    74784: gap of unknown length
            *    74785    87629: contig of 12845 bp in length
            *    87630    87729: gap of unknown length
            *    87730    88758: contig of 1029 bp in length
            *    88759    88858: gap of unknown length
            *    88859    96394: contig of 7536 bp in length
            *    96395    96494: gap of unknown length
            *    96495   126194: contig of 29700 bp in length
            *   126195   126294: gap of unknown length
            *   126295   136826: contig of 10532 bp in length
            *   136827   136926: gap of unknown length
            *   136927   138541: contig of 1615 bp in length
            *   138542   138641: gap of unknown length
            *   138642   142605: contig of 3964 bp in length
            *   142606   142705: gap of unknown length
            *   142706   156892: contig of 14187 bp in length
            *   156893   156992: gap of unknown length
            *   156993   158390: contig of 1398 bp in length.
FEATURES             Location/Qualifiers
     source          1. .158390
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="CTC-483I11"
                     /clone_lib="CalTech human BAC library C"
BASE COUNT    40327 a  39358 c  40000 g  37475 t   1230 others
ORIGIN


  Query Match             4.4%;  Score 43.2;  DB 60;  Length 158390;
  Best Local Similarity   45.0%;  Pred. No. 4.3;
  Matches  162;  Conservative    0;  Mismatches  198;  Indels    0;  Gaps    0;

Qy    124 ggccaaaaaacatcggtagtggcgacggcgaaaccggggtgtgtgtcctcgggcactgac 183
          |||   || ||  |||  | |||| ||| |      ||| |   |  | || ||      |
Db  87135 GGCATGAACACGCCGGAGGGGGCGGCGGTGGGGGGCGGGCGGCTTCCTTCAGGAGGGCGC 87076

Qy    184 gcggcgagtagcagtcataccaatacgacttctgccgctgctgcatcccctgcatcaccc 243
          ||     |       |   ||    || || |||| |  ||||   |  ||||  | |
Db  87075 GCCCGCCGAGCGCCGCTCTCCGCGGCGGCTGCTGCTGTGGCTGTCGCTGCTGCTGCCGCT 87016

Qy    244 cctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggctgagccg 303
          |    | ||| | | ||    |  || | | || |||| ||  | ||  | || |   ||
Db  87015 CTAACTGCAGTGGCGGCTGCTGTGGCTGTGGCTGCTGCTGCTCCCCGGGCCGCAGGCGCG 86956

Qy    304 cacgtggggagcaagatcattgctaatctagtgaatcagctggggattaatgtcacccaa 363
            || |  | |   | |  | |     || |||   | || |  |      | | |||
Db  86955 TCCGCGTCGTGGCCGCTGCTGGGATCGCTGGTGCCGCCGCCGCCGCCGCCAGGCGCCCGC 86896

Qy    364 aggagcgtcgtcagcactggagcgccggccacgacgaggtctacggcggtgacgtccacg 423
            |   || |   || | |    | | |||| |   | ||    |||    |      |
Db  86895 TCGCATGTGGCTCGCGCCGCGTCTCAGGCCGCCCGGCGGATCCCGGGCCCGCTCCAAAGC 86836

Qy    424 actaccgccccgcagcgaacaagtccatacgggcacaatggccgacctgtgacggctgga 483
```

```
              ||||||||| ||  |    ||   | |||        || || |   | ||  ||
Db  86835 CGCCCCGCCCCGCCCCGCCCCGCCCCCGCCCGGCGGGTCCCACGCCCCGCCCCCGCGTGA 86776


RESULT  13
EIMMAX/c
LOCUS       EIMMAX        767 bp    mRNA            INV       19-MAY-1993
DEFINITION  E.tenella antigen LPMC61 mRNA, partial cds.
ACCESSION   M30933
VERSION     M30933.1  GI:158895
KEYWORDS    antigen.
SOURCE      E.tenella (strain Wisconsin), mRNA to cDNA.
  ORGANISM  Eimeria tenella
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
            Eimeria.
REFERENCE   1  (bases 1 to 767)
  AUTHORS   Ko,C., Smith,C.K.II. and McDonell,M.
  TITLE     Identification and characterization of a target antigen of a
            monoclonal antibody directed against Eimeria tenella merozoites
  JOURNAL   Mol. Biochem. Parasitol. 41, 53-64 (1990)
  MEDLINE   90348718
COMMENT     Draft entry and computer-readable sequence for [Unpublished (1990)
            Synergen, INC., 1885 33rd St., Boulder,CO 80301] kindly submitted
            by C.S.Ko 12-DEC-1989.
FEATURES             Location/Qualifiers
     source          1. .767
                     /organism="Eimeria tenella"
                     /db_xref="taxon:5802"
     CDS             <1. .>767
                     /note="antigen LPCM61"
                     /codon_start=1
                     /protein_id="AAA29079.1"
                     /db_xref="GI:158896"
                     /translation="LRLLLKLLLLLLGQQKHWPERQQQQQPQPWLDRQQQQQQHNQQL
                     QKQQWPEGQRQQLWPEQQQQQWPEQHQQAQQQQQWPQQQPQMQQEQWPQQQPQVQQQQ
                     QWPQQQHRRQHGQQQQCMNSQQQLQQCGQQQQQQLQQQWSEQQQQQQQQQQWPEQPEQQ
                     QQQQWPEQQQQQWSDQNQQQQAQQWQAQQQQQWPQQQQQPQQQQQQQQQQDLGPDGVG
                     IVVPYLGSSPADFEVQNLGVSVLPRIKLPLPKFDNDDG"
     repeat_region   52. .634
                     /note="region of degenerate repeats"
BASE COUNT      201 a    240 c    249 g     77 t
ORIGIN


  Query Match             4.4%;  Score 42.8;  DB 6;  Length 767;
  Best Local Similarity   56.3%;  Pred. No. 25;
  Matches   80;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy    145 gcgacggcgaaaccggggtgtgtgtcctcgggcactgacgcggcgagtagcagtcatacc 204
          ||   ||| |   | |    |||  |   || ||| |  || ||   | ||  ||| ||
Db    205 GCTCAGGCCACTGCTGCTGCTGTTGCTCTGGCCACAGCTGCTGCCGCTGGCCCTCAGGCC 146

Qy    205 aatacgacttctgccgctgctgcatcccctgcatcacccctgtttcagcgccagccaag 264
          | | |  ||||||| |||| ||  |   ||||  |  |  |||||   |  | ||||| |
Db    145 ACTGCTGCTTCTGCAGCTGTTGGTTGTGCTGCTGCTGCTGTTGTCGATCTAGCCACG 86

Qy    265 gcagccgcgcctcctgccgcgg 286
          || || |   || |||| || |
Db     85 GCTGCGGTTGCTGCTGCTGCTG 64


RESULT  14
AP001081/c
LOCUS       AP001081   140812 bp    DNA             PLN       12-AUG-2000
DEFINITION  Oryza sativa genomic DNA, chromosome 1, clone:P0693B08.
ACCESSION   AP001081 BA000010
VERSION     AP001081.1  GI:6815077
KEYWORDS    .
SOURCE      Oryza sativa (cultivar:Nipponbare) DNA, clone:P0693B08.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
            Oryza.
REFERENCE   1  (bases 1 to 140812)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
            clone:P0693B08
  JOURNAL   Published Only in DataBase (2000) In press
REFERENCE   2  (bases 1 to 140812)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JAN-2000) to the DDBJ/EMBL/GenBank databases. Takuji
            Sasaki, National Institute of Agrobiological Resources, Rice Genome
            Research Program; Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     The orientation of the sequence is from SP6 to T7 of the PAC clone.
            Genes were predicted from the integrated results of the
            following:GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as
            SplicePredictor (October1998 version). The genomic sequence was
            searched against the non-redundant database NRP(PIR,SWISSPROT,
            GENPEPT, PDB) from MAFF DNAbank and the cDNA sequence database at
            RGP. Protein similarities of the cording regions were searched
            against NRP with BLASTP2.0. ESTs represent the identified cDNA
            sequences using BLASTN2.0 with the corresponding DDBJ accession no.
            and RGP clone ID
            The sequence of this clone has an overlap with P0667A10 clone,
            DDBJ:AP001073 at the 5' end and an overlap with P0424A08 clone,
            DDBJ:AP000837 at the 3' end. This clone starts at the position
            106,268 of P0667A10 and ends at the position 3,914 of P0424A08.
            Detailed information on assemble quality together with annotation
            of this entry at http://www.dna.affrc.go.jp:82/genomicdata/GenomeFi
            nished.html.
FEATURES             Location/Qualifiers
     source          1.. .140812
                     /organism="Oryza sativa"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:4530"
                     /chromosome="1"
                     /clone="P0693B08"
     CDS             join(3816. .4050,4056. .4279,4294. .4770)
                     /note="hypothetical protein"
                     /codon_start=1
                     /protein_id="BAA90364.1"
                     /db_xref="GI:6815078"
                     /translation="MPHFSSLSLSLTGGSAAAGDHGRRAGAEAAASGGWAQGWRRPGM
                     EEEAAGGGRGAEAGDGQWRRRQRAAGGREAGDGREAIVGGSQGAEVAQAPAGGEEPRP
                     KEAACGERGSSNLLLAALMRARVQRAWRARREPTHGGGGGEAAATRGAAAPTLRSGGH
                     RAPTTRRRSPPPIARPRSPGRPPPPSPAAVVLEPPPPPAIGSVPARRRPHLTAATLLC
                     RLRLRADVHAVCAAVHTVLVDRPSPRHPRRPPAPTARPRRARCAPAFSTPASRGPRLL
                     DHRPPPPPPPPVASLAPTRRPSPTSASRPLSPTHP"
     CDS             join(9496. .9723,10535. .11281)
                     /note="ESTs D24550(R2151),D24265(R1609),AU031848(R2151)
                     correspond to a region of the predicted gene.
                     Similar to cationic peroxidase isozyme 40K precursor
                     (D42065)"
                     /codon_start=1
                     /protein_id="BAA90365.1"
                     /db_xref="GI:6815079"
                     /translation="MDLAWWFAVAVVVCGLVGGGSAGLLETNPGLAYNFYQKSCPNVD
                     SIVRSVTWAQVAANPALPGRLLRLHFHDCFVQGCDASILLDNAGSEKTAGPNLSVGGY
                     EVIDAIKTQLEQACPGVVSCADIVALAARDAVSYQFKASLWQVETGRRDGPVSLASNT
                     GALPSPFAGFSTLLQSFANRGLNLTDLVALSGAHTIGKASCSSVTPRLYQGNTTSLDP
                     LLDSAYAKALMSSCPNPSPSSSTIDLDVATPLKFDSGYYANLQKKQGALASDAALTQN
                     AAAAQMVADLTNPIKFYAAFSMSMKKMGRIDVLTGSKGNIRKQCRSAS"
     CDS             join(13647. .13769,14188. .14391,14428. .14602,14993. .15468)
                     /note="hypothetical protein"
                     /codon_start=1
                     /protein_id="BAA90366.1"
                     /db_xref="GI:6815080"
                     /translation="MRVERVADLPVAMHQPTLLPVGSSSIATADTYEMNMQVGYRASR
                     AALLGTRPPGRGGRGGGGGSFDAGEGEGARLLRWRRVSCVSDKSGGGGESLLQASRIF
                     FLAILSPISASMYWSVSYTSRSGRAPPLGSAGPNPSFTTSLHTSIDLEYSRYQPRREL
                     VLVHSPAGPHMRRRDRRWLWAFAGAPRTGSTKTVRAQIIEQCTASPSCTLRKLALGAL
                     QLAEADHGAPRVRVFCIQPRGERLLHPQVDVRLHACRLYLGVPPPSVGVAAPCPGGLL
                     PSSAACLRREGKGEREMRKGENRGKREEADVGQADICNSCGSHADSAAT"
```

```
CDS             join(16052. .16064,16146. .16319,16355. .16935)
                /note="hypothetical protein"
                /codon_start=1
                /protein_id="BAA90367.1"
                /db_xref="GI:6815081"
                /translation="MGPTLGAAVEGERGDRWRSGDGGDRRGREVAGGGARVARKGKRG
                RRNGELGQLAAAAGEETVGRGEERRGAWHGAAAWERSGVAPTSSLFERTLTNLLAGIP
                ALLAGLLALLAGLPISFPSTPRASSPKRRSSTLEPPLCRSPVAATADHNDCLARTRKR
                EGTYTTGRIWRSSEPAAAPIDTLRSTPLPRRCAVRCFSPRLPRAQVGRARTGLPAASP
                SPSKPLCPTTRQPLPRSLAAHRLHPRIATPPAAGLSH"
CDS             18583. .20061
                /note="ESTs C71763(E0309),AU082059(E0309) correspond to a
                region of the predicted gene.
                Similar to Arabidopsis thaliana chromosome II BAC T20B5
                genomic sequence; hypothetical protein. (AC002409)"
                /codon_start=1
                /protein_id="BAA90368.1"
                /db_xref="GI:6815082"
                /translation="MAIAVGHCRELSPPGGGVRGRHCRLRSVTQLDCPVFDDHDLDLD
                RRERNGHAGLSPRRSLSSPCFSTVVPPAADRAEPARDDGEKMMPRVEIVAGGHARGVH
                ELIAEAAGAIATGTRLVPAQGGIGGALLLEDGRSVDHVAVIKPLLDDASSPSHGGGGG
                YASKAVLREVAAFLLDHDGFARVEPTALIKISRPAMPTTTASIQRFAAHECDAGELGP
                SRFSVASVHRIGSLDVRLLNIDRHAGNILVKKSPESECASGGSTLTPLDLVPIDHGLC
                LPEQLDDPYFEWLHWPQSSLPFSGAELEYVASLDPFRDAAMLRAELPSLTEAAIRILT
                LCTIFLQRAAAAGLCLADIGDMMTREFSAMEEGLSALESLCKNAYDSSTTTAGSSPRK
                HHHSDDSDDESTQFGMDDVPAGLPPHLFLLGGGGIAKSVSFSAAEQGAAGRGAARKRM
                SFKALSGDEWAAFLDRFEQLLPAALDAKKRAGLKLTRLGTSF"
LTR             complement(20793. .21346)
                /note="3' LTR"
CDS             join(21752. .21851,22552. .22676,22886. .23101)
                /note="hypothetical protein"
                /codon_start=1
                /protein_id="BAA90369.1"
                /db_xref="GI:6815083"
                /translation="MDLDKSTSTSVSLEKLQEEGALPGRGTMEREPGGLNSETVGRRI
                GQVMISGPTTASNVPVPLCENWAAEREAAINRPSHGRTARAHATRRHATAQARTAAAP
                AASRSRSRRWSRAPPCGQPPAWGQTSTASVAAWGRVQGARDAAS"
CDS             complement(join(23394. .23529,23558. .24018))
                /note="hypothetical protein"
                /codon_start=1
                /protein_id="BAA90370.1"
                /db_xref="GI:6815084"
                /translation="MGWPARRRTRRRRRRDLAQRLATASGGRSGGGGGGKEGDGARAV
                ECTGAGDCKTKKREEGTGVLYIGSRGLIVAGVAAISPAEWREVGRRERRDSKIESRPS
                RARTRAEEREWARGRRAQAATWATWPERRGSERGGGCGMQRRRLEVGDGPDREEGGGR
                LGWAVAFGWPSAEGERKRMGRRPIRKRKRREKKIKRRK"
CDS             complement(join(24161. .24471,24526. .24637))
                /note="hypothetical protein"
                /codon_start=1
                /protein_id="BAA90371.1"
                /db_xref="GI:6815085"
                /translation="MAVTRVAVHDEPGPRTGEGAVHHVALTWRDVAHTWQPHRTAAGS
                ARLARTGSSSRPGSNGVRRRYRGAAAADGGEWRRRDERRRRDATTRRRRYRRKREGEE
                GGGVLTEEVRPAWRETTVNGGETRQREVDKRRTGSGTT"
CDS             join(25452. .25542,26685. .26854,26925. .26983,27057. .27489,
                27514. .27825)
                /note="EST AU082058(C12976) corresponds to a region of the
                predicted gene.
                hypothetical protein"
                /codon_start=1
                /protein_id="BAA90372.1"
                /db_xref="GI:6815086"
                /translation="MERGGPYDGLCRFRIHLGTRGDCPLPAGDGVGPLVDHQAAASLK
                EDVAKEASDAAAAVATTSGVNVPKKGRKFSSVLGNRRKTPTPSASDASPPPPRRQRLV
                TLGEKAARAKAAQDRSGGNSSASPVVASTDVVVMTRSREATPNGPVGDLVPTQGSPAD
                ILTWEELQVEMGRILQAGARGIRHEIAEARAAAASSANERADKLARDLAEVREDLQKM
                RELVAGNERQRQGLKNRMSELENNLSEIRGSLRLARECGVTTTIPANPDEFLLTTSLA
                ELATAMEVIPSKHAARIGEETSSGIYTGTCNVLACVKLAHPELDLQKILYQGAASNAR
                KDVMEEIGDMGESILPLFEE"
CDS             complement(join(29314. .30354,33082. .35175))
                /note="Similar to GAG-POL precursor (AB030283)
                internal stop
                codon:complement(29869. .29871,30130. .30132,30295. .30297,
                34924. .34926,34960. .34962)
                probably inactive because stop codons are included in CDS"
                /codon_start=1
                /pseudo
LTR             complement(30355. .30714)
                /note="3' LTR"
CDS             complement(30656. .31903)
                /note="Similar to Zea mays 22-kDa alpha zein gene cluster;
                hypothetical protein. (AF031569)"
                /codon_start=1
                /protein_id="BAA90373.1"
                /db_xref="GI:6815087"
                /translation="MAEPAKAHDLSPSMSGDDGEPNPRRRARTPPPPPRQSPRREKAL
                ERAEGSATSTSTGGGEGRRDGERRLLVYGDGSTPQGALQAAGALLRHPPVAHDPESPA
                QRWLDDVAKLVMTARQRLDAGGRSSATKASGAATTGSASSRRRARRAAAAVRHSATTP
                SSTPSTREDLRGGPDARTSIERRRNGRRATHATEGASSSRVSPQHGRRNQPSVPPVGG
                VGCRAFVASLRNVRWPPRFRPTIAEKYDGSVNPAEFLQVYTTGIEAAGGDDRVMANFF
                PMALKGQARGWLMNLPPASVHSWEDLCQQFTMKFQGTYPRPGEEADLHAVQRRDDESL
                RSYIQRFCQVRNTIPCIPTHAVIYAFRGGVRHNRMLEKIASKEPQTTAELFQLADRVA
                RMKRWGKSPQKRGKWCMWYPLKI"
LTR             complement(31991. .32353)
                /note="5' LTR"
LTR             complement(38386. .38937)
                /note="5' LTR"
CDS             join(40089. .40184,40259. .40396,40823. .41078,41886. .41957,
                42532. .42671)
                /note="ESTs C91695(E31262),D40183(S1971) correspond to a
                region of the predicted gene.
                hypothetical protein"
                /codon_start=1
                /protein_id="BAA90374.1"
                /db_xref="GI:6815088"
                /translation="MAASAPPPAEPPPPPAAAAAAATNPPASTAPTEDGEVEEGATPP
                RTSALGTANWGTATLVGVFAGLLYGGSKEASASVSKDAEVMLKMGSTTDKREQYRLMR
                DAMEKRFIRVAKGSLVGGVRLGMFTATFFGIQNLLIEKRGVHDVFNIAGAGSATAAAF
                GLICSPMWRARNVLVGSALGAGVCFPLGWIQLKLAEKANLEMASSKPTDLVEEQGNQS
                GVGAAIERLGRSLKK"
CDS             complement(join(43206. .43424,44141. .44605,44699. .45774,
                45861. .46014,46110. .46206,46565. .46701))
                /note="Similar to glutamyl-tRNA synthetase (AF067773)"
                /codon_start=1
                /protein_id="BAA90375.1"
                /db_xref="GI:6815089"
                /translation="MEPKLAFPQDSPPLAILCAAKVAGVSLTLDPKLASGSAPTLHLG
```

```
  Query Match             4.4%;  Score 42.6;  DB 13;  Length 140812;
  Best Local Similarity   48.2%;  Pred. No. 6.1;
  Matches  120;  Conservative    0;  Mismatches  129;  Indels    0;  Gaps    0;

Qy    161 ggtgtgtgtcctcgggcactgacgcggcgagtagcagtcataccaatacgacttctgccg 220
          | |  | | || |||   | |  ||||      ||        ||   ||||  | || |
Db  72480 GTTCGGCGGCCCCGGCGGCCGCGGCGGACCCGCGCTCGGCCCTCACGGCGACGACGGCGG 72421

Qy    221 ctgctgcatcccctgcatcacccccctgtttcagcgccagccaaggcagccgcgcctcctg 280
          |  | ||  | || ||  |    | | |  || | | |  |  |||| || ||| |  | |
Db  72420 CACCGGCGGCACCGGCGGCTTTCTCGGCGTCGGAGGCGACGGAGGCCGCGGCGGCGACGG 72361

Qy    281 ccgcggcgcgatcggctgagccgcacgtggggagcaagatcattgctaatctagtgaatc 340
          | ||   |    | || ||| |    | || ||   || |||  ||  | || |  ||
Db  72360 CGGCCCAGGAGGCAGCGGAGGCCGCGGCGGCGACGGAGGTCAAGGCGGACCTGGCCAACG 72301

Qy    341 agctggggattaatgtcacccaaaggagcgtcgtcagcactggagcgccggccacgacga 400
            |  |||||     | | |     ||   |  | |    | ||  || |||| |||  |
Db  72300 GGGCGGGGACGGCGGGCCCGGCGGGGCAGGCGGGCCCGGCGGGCACGGCGGCGACGGTGG 72241

Qy    401 ggtctacgg 409
          || |  |||
Db  72240 GGACGGCGG 72232


RESULT  15
HSE4
LOCUS       HSE4         5880 bp    DNA             VRL       03-FEB-1999
```

```
DEFINITION  Equine herpesvirus 4 (EHV4) genome, thymidine kinase (TK) and
            glycoprotein H (gH) genes.
ACCESSION   D14486 D00683 D00684
VERSION     D14486.1  GI:221820
KEYWORDS    UL21 gene; UL24 gene; gH gene; glycoprotein; thymidine kinase.
SOURCE      Equine herpesvitus 4 ( strain 1942).
  ORGANISM  Equine herpesvirus 4
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Alphaherpesvirinae; Varicellovirus.
REFERENCE   1  (sites)
  AUTHORS   Nicolson,L., Cullinane,A.A. and Onions,D.E.
  TITLE     The nucleotide sequence of an equine herpesvirus 4 gene homologue
            of the herpes simplex virus 1 glycoprotein H gene
  JOURNAL   J. Gen. Virol. 71 (Pt 8), 1793-1800 (1990)
  MEDLINE   90362066
REFERENCE   2  (sites)
  AUTHORS   Nicolson,L., Cullinane,A.A. and Onions,D.E.
  TITLE     The nucleotide sequence of the equine herpesvirus 4 thymidine
            kinase gene
  JOURNAL   J. Gen. Virol. 71 (Pt 8), 1801-1805 (1990)
  MEDLINE   90362067
REFERENCE   3  (bases 1 to 5880)
  AUTHORS   Nicolson,L.
  JOURNAL   Unpublished (1993)
COMMENT     Submitted (10-SEP-1990) to DDBJ by:
            Lesley Nicolson
            Dept. Veterinary Pathology. University of Glasgow
            Vet School
            Bearsden Roda, Glasgow G61 1QH. Scotland
            UK.
            Phone: 041-339-8855
            Fax:   041-330-5733.
FEATURES             Location/Qualifiers
     source          1. .5880
                     /organism="Equine herpesvirus 4"
                     /db_xref="taxon:10331"
     polyA_signal    complement(99. .104)
                     /note="put. polyadenylation signal (UL24); putative"
     polyA_signal    complement(115. .120)
                     /note="put. polyadenylation signal (UL24); putative"
     CDS             complement(116. .934)
                     /note="putative;  ORF1 (UL24 homologue)"
                     /codon_start=1
                     /protein_id="BAA03377.1"
                     /db_xref="GI:221821"
                     /translation="MKRKQRLTARSRLRAGIRCHNRFYNAMVQDLASAKKNGVYGARL
                     APLFSELVPAETLKTAMGVSLAFEVNLGQRRPDCVCTVQFGHGSDAKGVCILIELKTC
                     RFSKNMNTASKNLQRKGGMRQLHDSCRLLAKTLPPGSGEILLAPVLVFVAQRGMRVLR
                     VTRLSPQVVYSNAAVLSCTISRLAEYSPPISERSTRRRCVTRRTNSKAFRAKTTTGSI
                     QPITQAKPAATAAVASLFSATAQANTTNAAVGYQPATISLANPLAWVASLFAPK"
     TATA_signal     861. .865
                     /note="put. TATA box (TK); putative"
     misc_feature    887
                     /note="pot. transcription initiation site; putative"
     CDS             949. .2007
                     /partial
                     /note="ORF2;  (TK)"
                     /codon_start=1
                     /protein_id="BAA03378.1"
                     /db_xref="GI:221822"
                     /translation="MAACVPPGEAPRSASGTPTRRQVTIVRIYLDGVYGIGKSTTGRV
                     MASAASGGSPTLYFPEPMAYWRTLFETDVISGIYDTQNRKQQGNLAVDDAALITAHYQ
                     SRFTTPYLILHDHTCTLFGGNSLQRGTQPDLTLVFDRHPVASTVCFPAARYLLGDMSM
                     CALMAMVATLPREPQGGNIVVTTLNVEEHIRRLRTRARIGEQIDITLIATLRNVYFML
                     VNTCHFLRSGRVWRDGWGELPTSCGAYKHRATQMDAFQERVSPELGDTLFALFKTQEL
                     LDDRGVILEVHAWALDALMLKLRNLNVFSADLSGTPRQCAAVVESLLPLMSSTLSDFD
                     SASALERAARTFNAEMGV"
     polyA_signal    2049. .2054
                     /note="put. polyadenylation signal (TK); putative"
     TATA_signal     2127. .2133
                     /note="putative"
     TATA_signal     2197. .2202
                     /note="putative"
     CDS             2225. .4792
                     /partial
                     /note="ORF3;  (gH)"
                     /codon_start=1
                     /protein_id="BAA03379.1"
                     /db_xref="GI:221823"
                     /translation="MSQPYLKIAILVAATIVSAIPVWTTPVSTSPPQQTKLHYVGNGT
                     WVHNNTFNVTRYDRITMEPVYNNNLSSTTFFVAISERNFRTVNTPLGASVFWILKSAL
                     NPPKHQPCIANVPEPGDPRGPCVNSTVSLFFNDNLEPFLMTKNLLEFEVLPDNYITGW
                     TFERSKTVATKGNPVGVVLSPPRTSPDVNNTIRDDGTPKQHLSIIDEHTTFVLDLQNF
                     TKTLTYISPFAAVWPITAFHAGITVMGCDTTQAIAYLGNGFMGLQISSVNNPPLEMIV
                     APNDVRARIVNRLPPRRRLEPPGPYAGPIYKVYVLSDGNFYLGHGMSKISREVAAYPE
                     ESLDYRYHLSLANLDTLAMLAELSSGKSKDVSYYLYRIIARLAVATFSLAEVIRLSDY
                     MLLQEAIDVDINLRLIVPLVMKYAAGGTADSSYTSSDVAMDQFEVAQAQIEKIVADIN
                     IENELRKPMYEHRSLLKSVYAYSRKPLPNAVSFANRLITAMYKEAIKDRITWNSTMRE
                     VLFFAVGAAAGSHVILTDGPDLGLHAHKDSSMFLSLNRNILLLCTAMCTASHAVSAGV
                     KLEEVMAGLIAGGVQFSLLEVFSPCMASARFDLAEEEHVLDLLSVIPPRLYTDLNTGL
                     EDDGTTIHSYGRSANGILNSRIAYNFDAVRVFTPELASCSTKLPKVLVVLPLASNRSY
                     VITRTAPNIGLTYSLDGVNIAKPIVISYITYGNCQVSRATIRSVYLDHPGHTQSCVYC
                     GSVFMRYMASGAIMDLIYIDDKDVELQLVAGENSTIPAFNPKLYTPSMNALLMFPNGT
                     VTLMSAFASYSAFKIPSTYLWASIGGLLLAILILYVIVKMLCGGVINNDYSLLLNSE"
     polyA_signal    4777. .4782
                     /note="put. polyadenylation signal(gH); putative"
     repeat_region   4893. .5315
                     /note="direct repeat region"
     rep_origin      5395. .5403
                     /note="put. replication origin; putative"
     rep_origin      5440. .5448
                     /note="put. replication origin; putative"
     CDS             complement(5755. .>5880)
                     /note="putative;  ORF4 (UL21 homologue)"
                     /codon_start=1
                     /protein_id="BAA03380.1"
                     /db_xref="GI:221824"
                     /translation="NVSLITAFDNLDLSRKGAYYLYYLLSERLKRGGVPVHVNRA"
BASE COUNT     1459 a   1377 c   1499 g   1545 t
ORIGIN


  Query Match             4.3%;  Score 41.6;  DB 59;  Length 5880;
  Best Local Similarity   44.1%;  Pred. No. 25;
  Matches  173;  Conservative    0;  Mismatches  219;  Indels    0;  Gaps    0;

Qy     145 gcgacggcgaaaccggggtgtgtgtcctcgggcactgacgcggcgagtagcagtcatacc 204
           ||| ||||||   | |     | | |   ||   |||  ||||||||          |
Db    4893 GCGGCGGCGAGGGCTGCTGCGGCGGCGAGGGCTGCTGCGGCGGCGAGGGCTGCGGCGGCG 4952

Qy     205 aatacgacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaag 264
             | | || | || || || ||     ||||  |  |  | |     ||  | ||   |
Db    4953 GCGAGGGCTGCGGCGGCGGCGGCGAGGGCTGCGGCGGCGGCGGCGAGGGCTGCGGCGGCG 5012

Qy     265 gcagccgcgcctcctgccgcggcgcgatcggctgagccgcacgtggggagcaagatcatt 324
           || ||   | || | || ||||||     |||||   ||   | || |||
Db    5013 GCGGCGAGGGCTGCGGCGGCGGCGGCGAGGGCTGCTGCGGCGGCGGCGAGGGCTGCTGCG 5072

Qy     325 gctaatctagtgaatcagctggggattaatgtcacccaaaggagcgtcgtcagcactgga 384
           ||          |  |    || |          |     |  |  |  ||  |||
Db    5073 GCGGCGGCGAGGGCTGCTGCGGCGGCGGCGAGGGCTGCTGCGGCGGCGGCGAGGGCTGCT 5132

Qy     385 gcgccggccacgacgaggtctacggcggtgacgtccacgactaccgccccgcagcgaaca 444
           |||  ||||  |||  |    || |||||| | ||    |  || | ||  ||  | |  |
Db    5133 GCGGCGGCGGCGAGGGCTGCTGCGGCGGCGGCGAGGGCTGCTGCGGCGGCGGCGAGGGCT 5192

Qy     445 agtccatacgggcacaatggccgacctgtgacggctggattggtggcagctaatagtggt 504
             | |    | |   |  |      ||| | |||| |    ||  || |||      |
Db    5193 GCTGCGGCGGCGGCGAGGGCTGCTGCTGCGGCGGCGGCGAGGGCTGCTGCTGCGGCGGCG 5252

Qy     505 gccagcgcggcctcgtctcccacagccgcggc 536
           || || || ||  || |  |  |    || ||
Db    5253 GCGAGGGCTGCTGCGGCGGCGGCGAGGGCTGC 5284
```

Search completed: September 21, 2001, 00:34:56
Job time: 29551 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: September 21, 2001, 02:41:41 ; Search time 330.77 Seconds
(without alignments)
1852.744 Million cell updates/sec

Title: US-09-138-735-1_COPY_1232_2207
Perfect score: 976
Sequence: 1 caggtacagcgtaacggctt.....tgtgaacgaggccattatgt 976

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0601:*
1: /cgn1_9/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /cgn1_9/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /cgn1_9/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /cgn1_9/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /cgn1_9/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /cgn1_9/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /cgn1_9/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /cgn1_9/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /cgn1_9/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /cgn1_9/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /cgn1_9/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /cgn1_9/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /cgn1_9/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /cgn1_9/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /cgn1_9/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /cgn1_9/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /cgn1_9/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /cgn1_9/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	976	100.0	3402	20	AAX84092	T. cruzi PTc40 cod
	2	972.8	99.7	3402	17	AAT27310	Trypanosoma cruzi
c	3	40.4	4.1	477	21	AAZ44307	Human SCA7 genomic
c	4	40.4	4.1	567	21	AAA29550	HIV codon altered
	5	40	4.1	10732	21	AAA10594	Gene encoding a su
	6	39.2	4.0	2790	19	AAV18480	BOP1 cDNA. Mus sp
c	7	38.2	3.9	2943	17	AAT16480	SA8 virus gB glyco
	8	38	3.9	1037	21	AAA59242	Exons E, C and A o
	9	38	3.9	1159	21	AAA59240	An EcoRI fragment
	10	38	3.9	1472	21	AAA59241	Exons D, C, B and
	11	37.8	3.9	371	20	AAZ19383	M. tuberculosis an
	12	37.8	3.9	371	20	AAZ19171	M. tuberculosis re
	13	37.6	3.9	2115	20	AAZ32263	Human dunce-like p
	14	37.6	3.9	2115	21	AAA88198	pPDE39 human dunce
	15	37.6	3.9	72750	21	AAA81468	N. meningitidis pa
c	16	37.6	3.9	349980	21	AAF21544	Neisseria meningit
	17	37.2	3.8	921	20	AAX33558	Rice beta-glucanas
	18	37.2	3.8	996	20	AAX33568	Rice Gns6 CDS. Or
	19	37.2	3.8	2169	20	AAX33534	Rice beta-glucanas
	20	36.8	3.8	3737	21	AAA64656	Partial sequence c
	21	36.8	3.8	3776	17	AAT42903	TRP-1 protein codi
c	22	36.6	3.8	203	19	AAV30271	Glutamine rich reg
c	23	36.6	3.8	203	19	AAV17226	SCA2 gene CAG repe
	24	36.6	3.8	913	17	AAT30253	Cotton fibre cell-
	25	36.6	3.8	913	17	AAT13034	Cotton fibre-speci
	26	36.6	3.8	913	17	AAT07199	Cotton fiber-speci
	27	36.6	3.8	913	18	AAT70041	Cotton fibre speci
	28	36.6	3.8	913	18	AAT62610	Cotton fibre speci
	29	36.6	3.8	913	21	AAZ35545	cDNA sequence a co
	30	36.6	3.8	1966	20	AAX61220	Mouse DNA demethyl
	31	36.6	3.8	1984	17	AAT13030	Cotton fibre-speci
	32	36.6	3.8	1985	17	AAT30250	Cotton fibre clone
	33	36.6	3.8	1985	18	AAT70036	Cotton H6 gene and
c	34	36.6	3.8	3324	20	AAX60262	Nucleic acid seque
c	35	36.4	3.7	2415	18	AAT86757	cDNA of the M3/6 g
c	36	36.4	3.7	2453	18	AAT86758	cDNA of the M3/6 g
c	37	36.2	3.7	336	20	AAX88093	Synthetic antifree
c	38	36.2	3.7	633	11	AAQ03322	Genomic Eimeria te
	39	36.2	3.7	862	21	AAF09682	Fusarium venenatum
	40	36.2	3.7	3201	21	AAA14955	DNA encoding a PDE
c	41	36	3.7	795	19	AAV55830	FLGA insert stabil
c	42	36	3.7	799	19	AAV55831	Nucleotide sequenc
	43	36	3.7	1448	21	AAF21931	Human breast and o
c	44	36	3.7	1925	20	AAX90924	Epstein Barr Virus
c	45	36	3.7	1926	21	AAA50254	Epstein Barr virus

ALIGNMENTS

RESULT 1
AAX84092
ID AAX84092 standard; cDNA; 3402 BP.
XX
AC AAX84092;
XX
DT 27-AUG-1999 (first entry)
XX
DE T. cruzi PTc40 coding sequence.
XX
KW PTc40; Tc40; infection; diagnosis; immune complex; antigenic determinant;
KW therapy; antibody; ds.
XX
OS Trypanosoma cruzi.
XX
PN WO9929867-A1.
XX
PD 17-JUN-1999.
XX
PF 10-DEC-1998; 98WO-IB01987.
XX
PR 10-DEC-1997; 97US-0988242.
XX
PA (INMR) BIO MERIEUX.
XX
PI Jolivet M, Lesenechal M, Mandrand B, Paranhos-Baccala G;
XX
DR WPI; 1999-394978/33.
DR P-PSDB; AAY22124.
XX
PT New Trypanosoma cruzi antigen
XX
PS Claim 1; Page 52-56; 65pp; English.
XX

```
CC   This sequence encodes the Trypanosoma cruzi PTc40 protein of the
CC   invention, and is designated Tc40. The PTc40 antigenic determinant is
CC   useful as a reagent for detection and/or monitoring of Trypanosoma cruzi
CC   infection from samples including blood serum or plasma, urine, saliva, or
CC   tears, by contacting with the sample and detecting an immune complex. The
CC   PTc40 antigenic determinant, the vector, expression cassette, cell or
CC   antibody are useful for treatment or prevention (vaccine) of a
CC   Trypanosoma cruzi infection in a man or animal. Current Trypanosoma cruzi
CC   antigens are obtained from protein fractions of the noninfectious stage
CC   of the parasite, and these do not allow sufficient production of antigens
CC   for use in reliable serological diagnostic tests. The strain to strain
CC   polymorphism reduces reliability of the tests.
XX
SQ   Sequence 3402 BP; 888 A; 821 C; 956 G; 737 T; 0 other;


  Query Match             100.0%;  Score 976;  DB 20;  Length 3402;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-257;
  Matches  976;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 caggtacagcgtaacggcttttgcttcaatcgtacagccgacggtagctgcgtcctggct 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1232 caggtacagcgtaacggcttttgcttcaatcgtacagccgacggtagctgcgtcctggct 1291

Qy      61 gacatgtcgattcgattgacgatcttccatctccggtcctcccgcagggaagaacagcag 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1292 gacatgtcgattcgattgacgatcttccatctccggtcctcccgcagggaagaacagcag 1351

Qy     121 ccaggccaaaaaacatcggtagtggcgacggcgaaaccggggtgtgtgtcctcgggcact 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1352 ccaggccaaaaaacatcggtagtggcgacggcgaaaccggggtgtgtgtcctcgggcact 1411

Qy     181 gacgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatcccctgcatca 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1412 gacgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatcccctgcatca 1471

Qy     241 ccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggctgag 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1472 ccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggctgag 1531

Qy     301 ccgcacgtggggagcaagatcattgctaatctagtgaatcagctggggattaatgtcacc 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1532 ccgcacgtggggagcaagatcattgctaatctagtgaatcagctggggattaatgtcacc 1591

Qy     361 caaaggagcgtcgtcagcactggagcgccggccacgacgaggtctacggcggtgacgtcc 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1592 caaaggagcgtcgtcagcactggagcgccggccacgacgaggtctacggcggtgacgtcc 1651

Qy     421 acgactaccgccccgcagcgaacaagtccatacgggcacaatggccgacctgtgacggct 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1652 acgactaccgccccgcagcgaacaagtccatacgggcacaatggccgacctgtgacggct 1711

Qy     481 ggattggtggcagctaatagtggtgccagcgcggcctcgtctcccacagccgcggcgaaa 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1712 ggattggtggcagctaatagtggtgccagcgcggcctcgtctcccacagccgcggcgaaa 1771

Qy     541 ccaacaggagaagaaaaggcctccgcggcatgtgaaacgagctccgtggcgataaatgcg 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1772 ccaacaggagaagaaaaggcctccgcggcatgtgaaacgagctccgtggcgataaatgcg 1831

Qy     601 acacgcccggcgcttcacaacgcctctctcccgcaggcgccaacggatggcgttttggcg 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1832 acacgcccggcgcttcacaacgcctctctcccgcaggcgccaacggatggcgttttggcg 1891

Qy     661 gcagcagtataccagtcggagggcgaggttcatcagtcgctggagcggctggagtccgtc 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1892 gcagcagtataccagtcggagggcgaggttcatcagtcgctggagcggctggagtccgtc 1951

Qy     721 ataaccaacacgtctcgggttctgaagttgctccctgacaccattcgaagagaccatgaa 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1952 ataaccaacacgtctcgggttctgaagttgctccctgacaccattcgaagagaccatgaa 2011

Qy     781 caacttctgaatctgggtttagaggcacagatgacagagctgcagcagagccgtccaaca 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2012 caacttctgaatctgggtttagaggcacagatgacagagctgcagcagagccgtccaaca 2071

Qy     841 ccgcaaacacagccgagagacacaagctccgcgaaatcatccgtgtttgagacgtacacc 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2072 ccgcaaacacagccgagagacacaagctccgcgaaatcatccgtgtttgagacgtacacc 2131

Qy     901 cttgttctcattgcggattccctctctcgcaacatcacgaaggggtgaagcgtggtgtg 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2132 cttgttctcattgcggattccctctctcgcaacatcacgaaggggtgaagcgtggtgtg 2191

Qy     961 aacgaggccattatgt 976
           ||||||||||||||||
Db    2192 aacgaggccattatgt 2207


RESULT   2
AAT27310
ID   AAT27310 standard; cDNA; 3402 BP.
XX
AC   AAT27310;
XX
DT   26-NOV-1996  (first entry)
XX
DE   Trypanosoma cruzi epimastigotic PTc100t antigen gene.
XX
KW   Antigen; Trypanosoma cruzi; epimastigote; serum; Chagas disease; probe;
KW   Primer; PCR; polymerase chain reaction; amplification; antibody; ds.
XX
OS   Trypanosoma cruzi.
XX
FH   Key             Location/Qualifiers
FT   CDS             266..3013
FT                   /*tag=  a
FT                   /product= PTc100t epimastogotic antigen
XX
PN   FR2723589-A1.
XX
PD   16-FEB-1996.
XX
PF   12-AUG-1994;   94FR-0010132.
XX
PR   12-AUG-1994;   94FR-0010132.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Jolivet M,  Lesenechal M,  Paranhos-Baccala G;
XX
DR   WPI; 1996-190287/20.
DR   P-PSDB; AAR91615.
XX
PT   New nucleic acid encoding Trypanosoma cruzi epimastigotic antigen -
PT   useful for diagnosis, monitoring and therapy of Chagas disease
XX
PS   Claim 1; Page 24-26; 55pp; French.
XX
CC   This is the nucleotide sequence encoding a novel isolated antigenic
CC   protein from Trypanosoma cruzi epimastigotes, designated PTc100t.
CC   The clone Tc50 was isolated from a T.cruzi genomic expression library in
CC   lambda gt11, using a mixture of sera from patients with Chagas disease.
CC   Clone Tc50 contained an 594 bp insert corresp. to nucleotides 1232-1825
CC   of this sequence.  The Tc50 sequence was subsequently used to probe a
CC   Southern blot of restriction enzyme digested T.cruzi DNA and also screen
CC   a lambda gt10 library to isolate a 1041 bp EcoRI fragment corresp. to
CC   nucleotides 1403-2443 of PTc100t.  Primers (AAT27311-5) were synthesised
CC   based on the sequences of the 594 and 1041 bp fragments and used to
CC   amplify the PTc100t clone as 3 fragments from cDNA derived from mRNA
CC   purified from T.cruzi epimastigotes.  The protein or antibodies raised
CC   against it can be used in the detection and monitoring of T.cruzi
CC   infection i.e. Chagas disease.
XX
SQ   Sequence 3402 BP; 889 A; 818 C; 958 G; 737 T; 0 other;
```

```
  Query Match             99.7%;  Score 972.8;  DB 17;  Length 3402;
  Best Local Similarity   99.8%;  Pred. No. 1.7e-256;
  Matches  974;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 caggtacagcgtaacggcttttgcttcaatcgtacagccgacggtagctgcgtcctggct 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1232 caggtacagcgtaacggcttttgcttcaatcgtacagccgacggtagctgcgtcctggct 1291

Qy     61 gacatgtcgattcgattgacgatcttccatctccggtcctcccgcagggaagaacagcag 120
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1292 gacatgtcgaatcgattgacgatcttccatctccggtcctcccgcagggaagaacagcag 1351

Qy    121 ccaggccaaaaaacatcggtagtggcgacggcgaaaccggggtgtgtgtcctcgggcact 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1352 ccaggccaaaaaacatcggtagtggcgacggcgaaaccggggtgtgtgtcctcgggcact 1411

Qy    181 gacgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatcccctgcatca 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1412 gacgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatcccctgcatca 1471

Qy    241 ccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggctgag 300
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   1472 ccccctgtttcagcgccagcgaaggcagccgcgcctcctgccgcggcgcgatcggctgag 1531

Qy    301 ccgcacgtggggagcaagatcattgctaatctagtgaatcagctggggattaatgtcacc 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1532 ccgcacgtggggagcaagatcattgctaatctagtgaatcagctggggattaatgtcacc 1591

Qy    361 caaaggagcgtcgtcagcactggagcgccggccacgacgaggtctacggcggtgacgtcc 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1592 caaaggagcgtcgtcagcactggagcgccggccacgacgaggtctacggcggtgacgtcc 1651

Qy    421 acgactaccgccccgcagcgaacaagtccatacgggcacaatggccgacctgtgacggct 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1652 acgactaccgccccgcagcgaacaagtccatacgggcacaatggccgacctgtgacggct 1711

Qy    481 ggattggtggcagctaatagtggtgccagcgcggcctcgtctcccacagccgcggcgaaa 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1712 ggattggtggcagctaatagtggtgccagcgcggcctcgtctcccacagccgcggcgaaa 1771

Qy    541 ccaacaggagaagaaaaggcctccgcggcatgtgaaacgagctccgtggcgataaatgcg 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1772 ccaacaggagaagaaaaggcctccgcggcatgtgaaacgagctccgtggcgataaatgcg 1831

Qy    601 acacgcccggcgcttcacaacgcctctctcccgcaggcgccaacggatggcgttttggcg 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1832 acacgcccggcgcttcacaacgcctctctcccgcaggcgccaacggatggcgttttggcg 1891

Qy    661 gcagcagtataccagtcggagggcgaggttcatcagtcgctggagcggctggagtccgtc 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1892 gcagcagtataccagtcggagggcgaggttcatcagtcgctggagcggctggagtccgtc 1951

Qy    721 ataaccaacacgtctcgggttctgaagttgctccctgacaccattcgaagagaccatgaa 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1952 ataaccaacacgtctcgggttctgaagttgctccctgacaccattcgaagagaccatgaa 2011

Qy    781 caacttctgaatctgggtttagaggcacagatgacagagctgcagcagagccgtccaaca 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2012 caacttctgaatctgggtttagaggcacagatgacagagctgcagcagagccgtccaaca 2071

Qy    841 ccgcaaacacagccgagagacacaagctccgcgaaatcatccgtgtttgagacgtacacc 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2072 ccgcaaacacagccgagagacacaagctccgcgaaatcatccgtgtttgagacgtacacc 2131

Qy    901 cttgttctcattgcggattccctctctcgcaacatcacgaaggggtgaagcgtggtgtg 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2132 cttgttctcattgcggattccctctctcgcaacatcacgaaggggtgaagcgtggtgtg 2191

Qy    961 aacgaggccattatgt 976
          ||||||||||||||||
Db   2192 aacgaggccattatgt 2207
```

```
RESULT   3
AAZ44307/c
ID   AAZ44307 standard; DNA; 477 BP.
XX
AC   AAZ44307;
XX
DT   04-APR-2000  (first entry)
XX
DE   Human SCA7 genomic DNA fragment 1.
XX
KW   SCA7; human; spinocerebellar ataxia type 7; SCA1; SCA2; SCA3; SCA6;
KW   repeat expansion detection; RED analysis; detection; ds.
XX
OS   Homo sapiens.
XX
PN   CA2245310-A.
XX
PD   19-FEB-1999.
XX
PF   19-AUG-1998;   98CA-2245310.
XX
PR   19-AUG-1997;   97US-0056170.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Koob MD,  Ranum LP;
XX
DR   WPI; 2000-098181/09.
XX
PT   Identifying individuals at risk of developing spinocerebellar ataxia
PT   type 7 by analyzing trinucleotide repeat regions of spinocerebellar
PT   ataxia type 7 gene -
XX
PS   Claim 36; Page 41; 66pp; English.
XX
CC   This invention describes a novel method for identifying individuals at
CC   risk for developing spinocerebellar ataxia type 7 (SCA7). The method
CC   comprises analyzing the CAG repeat region of a SCA7 gene to detect CAG
CC   repeats, where individuals at risk have at least 30 CAG repeats and
CC   those not at risk have less than 19 CAG repeats. The method is useful
CC   for identifying individuals at risk of developing SCA7 and also those
CC   at risk of developing SCA1, 2, 3 or 6. The use of genomic DNA in the
CC   repeat expansion detection (RED) analysis allows isolation of any
CC   potential trinucleotide repeat expansion regardless of the expression
CC   pattern. Utilization of different oligonucleotides in the RED assay
CC   allows any of the possible trinucleotide repeats to be detected, and
CC   the cycled nature of the reaction makes it extremely sensitive. This
CC   sequence represents a genomic DNA fragment of the human SCA7 gene
CC   which is described in the method of the invention.
XX
SQ   Sequence 477 BP; 99 A; 164 C; 172 G; 42 T; 0 other;
```

```
  Query Match             4.1%;  Score 40.4;  DB 21;  Length 477;
  Best Local Similarity   53.9%;  Pred. No. 0.11;
  Matches   83;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy    145 gcgacggcgaaaccggggtgtgtgtcctcgggcactgacgcggcgagtagcagtcatacc 204
          | | |||||    ||||   ||   || | ||  |||  | ||||        |  | |
Db    334 GTGGCGGCGGTGGCGGGTGCTGCTGCCGCTGGGGCTGCGGAGGCGGCGGCTGCTGCTGCT 275

Qy    205 aatacgacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaag 264
            | |  || |||| |||||||||  |  ||||  |  |  ||| | | ||  | ||
Db    274 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 215

Qy    265 gcagccgcgcctcctgccgcggcgcgatcggctg 298
          || || ||  || |||| || ||     | ||||
Db    214 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 181
```

```
RESULT   4
AAA29550/c
```

```
ID   AAA29550 standard; DNA; 567 BP.
XX
AC   AAA29550;
XX
DT   09-AUG-2000  (first entry)
XX
DE   HIV codon altered env nucleotide sequence #2.
XX
KW   Erythropoietin; EPO; G-CSF; granulocyte colony stimulating factor;
KW   wobble; codon altered gene; shuffling; modification; vaccine;
KW   insulin; peptide hormone; growth factor; cytokine; interferon;
KW   interleukin; leukaemia inhibitory factor; oncostatin M;
KW   transcription activator; expression activator; infectious organism; ds.
XX
OS   Human immunodeficiency virus type 1.
OS   Synthetic.
XX
PN   WO200018906-A2.
XX
PD   06-APR-2000.
XX
PF   28-SEP-1999;   99WO-US22588.
XX
PR   29-SEP-1998;   98US-0102362.
PR   29-JAN-1999;   99US-0117729.
PR   05-FEB-1999;   99US-0118813.
PR   24-JUN-1999;   99US-0141049.
XX
PA   (MAXY-) MAXYGEN INC.
XX
PI   Patten PA,  Liu L,  Stemmer WPC;
XX
DR   WPI; 2000-303449/26.
XX
PT   Novel methods for recombining codon-altered libraries of nucleic acids
PT   used to produce new proteins and new vectors with reduced rates of
PT   reversion to wild type  -
XX
PS   Example; Fig 18A; 92pp; English.
XX
CC   A method has been developed of making codon altered nucleic acids (NAs),
CC   comprising providing a NA sequence (NA1) which encodes a polypeptide
CC   (P1), providing codon altered NA sequences, each encoding P1 or a
CC   modified form of it, and recombining the codon altered NA sequences to
CC   produce a target codon altered NA which encodes a second protein.
CC   The method of the invention can be used for recombining codon-altered
CC   libraries of nucleic acids to produce new proteins, which have
CC   improvements in a desirable characteristic. Target nucleic acids
CC   include those coding for therapeutic proteins such as erythropoietin
CC   (EPO), insulin, peptide hormones, growth factors, cytokines, interferons,
CC   interleukins, leukaemia inhibitory factor, and oncostatin M, as well as
CC   transcription and expression activators and proteins from infectious
CC   organisms for use as vaccines. The method can also be used to produce
CC   attenuated viruses which have reduced rates of reversion to wild type.
CC   The present sequence represents an HIV codon altered env nucleotide
CC   sequence, which is used in an example from the present invention.
XX
SQ   Sequence 567 BP; 32 A; 129 C; 222 G; 184 T; 0 other;


  Query Match             4.1%;  Score 40.4;  DB 21;  Length 567;
  Best Local Similarity   49.1%;  Pred. No. 0.12;
  Matches  107;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy     386 cgccggccacgacgaggtctacggcggtgacgtccacgactaccgccccgcagcgaaacaa 445
           || || | ||||||| | | ||| ||  |||| | ||||| ||  |  ||  |   || |
Db     477 CGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGA 418

Qy     446 gtccatacgggcacaatggccgacctgtgacggctggattggtggcagctaatagtggtg 505
           |      |   |  |    | | |   | || | |   |  |||||       |  |
Db     417 CGACGACGACGACGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 358

Qy     506 ccagcgcggcctcgtctcccacagccgcggcgaaaccaacaggagaagaaaaggcctccg 565
```

```
           |    || ||  |  |  |  |||| || ||   | || ||| || || |   ||  | |
Db     357 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 298

Qy     566 cggcatgtgaaacgagctccgtggcgataaatgcgaca 603
           | |||   | ||| |   |     | | ||   | |||
Db     297 CAGCAGCAGCAACAACAACAACAACAACAACAACAACA 260


RESULT   5
AAA10594
ID   AAA10594 standard; DNA; 10732 BP.
XX
AC   AAA10594;
XX
DT   29-JUN-2000  (first entry)
XX
DE   Gene encoding a subunit of cellulose synthase.
XX
KW   Cellulose synthase; cellulose production; increase yield; ds.
XX
OS   Vigna angularis.
XX
PN   JP2000060568-A.
XX
PD   29-FEB-2000.
XX
PF   26-AUG-1998;   98JP-0239998.
XX
PR   26-AUG-1998;   98JP-0239998.
XX
PA   (MIZU/) MIZUNO K.
PA   (OJIP ) OJI PAPER CO.
XX
DR   WPI; 2000-342371/30.
DR   P-PSDB; AAY85179.
XX
PT   A gene encoding a cellulose synthetic equipment - for the improvement
PT   in the amount of cellulose synthesised in a plant body
XX
PS   Claim 2; Page 14-21; 32pp; Japanese.
XX
CC   This sequence represents a gene encoding a subunit of the cellulose
CC   synthase complex of Vigna angularis. The invention relates to subunits of
CC   cellulose synthetic equipment, that can be used to increase the amount of
CC   cellulose synthesised by a plant. The proteins and genes encoding them
CC   can also be used to improve the properties of the cellulose being
CC   produced by a plant.
XX
SQ   Sequence 10732 BP; 3149 A; 1212 C; 2074 G; 2046 T; 2251 other;


  Query Match             4.1%;  Score 40;  DB 21;  Length 10732;
  Best Local Similarity   13.0%;  Pred. No. 0.45;
  Matches   68;  Conservative  218;  Mismatches  235;  Indels    1;  Gaps    1;

Qy      92 tccggtcctcccgcagggaagaacagcagccaggccaaaaaacatcggtagtggcgacgg 151
           |: || |:   :   | :  ::|: |:: : |:|::|     :   :| | |    :
Db    9423 thrggncysaavagggasnysvaysgsrasnargysasnggnhsrargvamtthrargth 9482

Qy     152 cgaaaccggggtgtgtgtcctcgggcactgacgcggcgagtagcagtcataccaatacga 211
           | ::: :|: | | ::::    :::: |:: ::: : : | ::   : :: |
Db    9483 raasryshsgsrgngashsysgyhssrvaghshssrasnvaaasrhsghsrhaasrsras 9542

Qy     212 cttctgccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaaggcagccg 271
           : : : ::: :::: ::    :    |:   : | :::  |:   ::  :|  :|::
Db    9543 aagysrasasysthrsrcysaagsrtraasraasrgashshargsrgysysysgrsrcys 9602

Qy     272 cgcctcctgccgcggcgcgatcggctgagccgcacgtggggagcaagatcattgctaatc 331
           :::    :  :: : : : |  : : |:  :: |:   :::||: |  :    :  :| :
Db    9603 yssrvagyaasrasasnmtasashvagmtgysaavavasrvagysaathrgnsrrsraas 9662

Qy     332 tagtgaatcagctggggattaatgtcacccaaaggagcgtcgtcagcactggagcgccgg 391
           :  :||::|   :::: |::  |::::  :  ::| |:::  :  |::::::|  :
```

```
Db    9663 rysaavathrgasnsrsrgthrgythrysgsrthrrgggyysgrvarashssrgnthrth 9722

Qy     392 ccacgacgaggtctacggcggtgacgtccacgactaccgccccgcagcgaacaagtccat 451
           :: :::|: : |:   : |:: ::::: :  : :  : :  :: :: :  :   : :
Db    9723 rgthrthrrgvavaasmtgrmtsrashssrassryssrysasnthrcyssrashysgyth 9782

Qy     452 acgggcacaatggccgacctgtgacggctggattggtggcagctaatagtggtgccagcg 511
           :: :|:     :: :  : :: |  : : :  :  :::|:: :  : |:::    :|::
Db    9783 rvargytrgnasvavaysthrggnasnhsvathrhsysgyrasashsasysvaaaarghs 9842

Qy     512 cggcctcgtctcccacagccgcggcgaaacca-acaggagaagaaaaggcctccgcggca 570
           :  :  :  :: ::|: ::   ||::::|::|  ::| |: ::   : :|::::| :::
Db    9843 vaasnashashashassrsrarggysrvahsasthrgnasnrrgncyshscyssrcyssr 9902

Qy     571 tgtgaaacgagctccgtggcgataaatgcgacacgcccggcg 612
            : : :: :| :  ::  |:|| || ::: :| |  ::| |
Db    9903 assrsrvavaasnrsraagygasnaaassrysarggrsrgng 9944
```

RESULT 6
AAV18480
ID AAV18480 standard; cDNA to mRNA; 2790 BP.
XX
AC AAV18480;
XX
DT 18-AUG-1998 (first entry)
XX
DE BOP1 cDNA.
XX
KW Corticotroph; pituitary; BOP1; tumour suppressor; constitutive;
KW inducible; alzheimer's disease; nuclear transcription factor; apoptosis;
KW cell cycle; neuronal disorder; ss.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT CDS 542..2545
FT /*tag= a
FT /product= "BOP1 protein"
XX
PN WO9813489-A1.
XX
PD 02-APR-1998.
XX
PF 22-SEP-1997; 97WO-EP05198.
XX
PR 23-SEP-1996; 96US-0718661.
XX
PA (CNRS) CENT NAT RECH SCI.
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI Journot L, Spengler D;
XX
DR WPI; 1998-230701/20.
DR P-PSDB; AAW48760.
XX
PT New isolated tumour suppressor gene - useful for developing products
PT for use in diagnosis and treatment of tumour(s) or neuronal
PT disorder(s)
XX
PS Claim 1; Pages 72-76; 118pp; English.
XX
CC The present sequence represents the BOP1 cDNA isolated from the
CC mouse corticotroph pituitary tumour cell line AtT-20 cDNA library.
CC The protein encoded by the BOP1 cDNA displays a tumour suppressing
CC activity when it was constitutively and inducibly expressed in
CC tumour cells. The BOP1 cDNA and the protein it encodes are claimed
CC to be useful in the preparation of therapeutic compositions, useful
CC for treating, preventing or delaying the recurrence of a tumour or
CC neuronal disorders, e.g. genetic diseases or acquired degenerative
CC encephalopathies such as Alzheimer's disease. The BOP1 protein is
CC also claimed to be able to induce apoptosis resulting in inhibition
CC of tumour cell growth, to suppress tumour formation, to induce G1
CC arrest of the cell cycle and to act as nuclear transcription factor.
XX
SQ Sequence 2790 BP; 667 A; 783 C; 714 G; 626 T; 0 other;

```
  Query Match              4.0%;  Score 39.2;  DB 19;  Length 2790;
  Best Local Similarity   54.1%;  Pred. No. 0.45;
  Matches   80;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy     137 cggtagtggcgacggcgaaaccggggtgtgtgtcctcgggcactgacgcggcgagtagca 196
           | | ||  || || || |   | |         || | | ||||| | | || | |  ||
Db    2114 cagcagcagccaccgccactgccgccgccaccaccgctgccactgccacagccactgcca 2173

Qy     197 gtcataccaatacgacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgc 256
           ||   ||    || ||  | ||| | | |||| || | |  |||    | | | |||| |
Db    2174 gtgccacagccactaccacagccacagatgcagccacagtttcagttgcagatccagccc 2233

Qy     257 cagccaaggcagccgcgcctcctgccgc 284
           |||     || || |  || |||||||
Db    2234 cagatgcagctaccacagctgctgccgc 2261
```

RESULT 7
AAT16480/c
ID AAT16480 standard; DNA; 2943 BP.
XX
AC AAT16480;
XX
DT 11-MAY-1996 (first entry)
XX
DE SA8 virus gB glycoprotein coding sequence.
XX
KW Herpes simian monkey SA8 virus gB glycoprotein; immunoassay;
KW diagnosis; herpes B virus; ss.
XX
OS Herpes simian monkey SA8 virus.
XX
FH Key Location/Qualifiers
FT CDS 87..2744
FT /*tag= a
FT /product= gB_glycoprotein
XX
PN US5487969-A.
XX
PD 30-JAN-1996.
XX
PF 01-APR-1993; 93US-0042747.
XX
PR 01-APR-1993; 93US-0042747.
XX
PA (SWBI-) SOUTHWEST FOUND BIOMEDICAL RES.
XX
PI Black D, Eberle R, Hilliard J, Scinicariello F;
XX
DR WPI; 1996-105220/11.
DR P-PSDB; AAR92747.
XX
PT Detection of herpes B virus by PCR amplification of sample DNA - to
PT detect a specific herpes simian monkey B virus DNA segment.
XX
PS Disclosure; Column 23-30; 22pp; English.
XX
CC This sequence encodes the herpes simian monkey SA8 virus gB
CC glycoprotein. Such sequences have been used to develop synthetic
CC DNA primers which have homologous sequences of conserved regions
CC which flank a divergent region of the gB glycoprotein gene. In
CC addition, proteins such as the gB glycoprotein have potential use in
CC the development of serological immunoassays. One approach is to
CC synthesize peptides which, based on the properties of the predicted
CC protein sequence, are likely to be immunologically active. Such
CC peptides can be used as substrate antigens in immunoassays to detect
CC serum antibodies which recognize this specific peptide sequence.
CC Synthetic peptides may also be used to produce antibodies against

```
CC   specific regions of the gB glycoprotein which are unique to one
CC   virus.  These can then be used to develop virus-specific
CC   immunoassays for differentiation of SA8 virus from other primate
CC   alpha-herpes viruses and for identification of antibodies directed
CC   against SA8 virus in primate serum samples.
XX
SQ   Sequence 2943 BP; 462 A; 1103 C; 983 G; 395 T; 0 other;


  Query Match             3.9%;  Score 38.2;  DB 17;  Length 2943;
  Best Local Similarity   47.7%;  Pred. No. 0.87;
  Matches  112;  Conservative    0;  Mismatches  123;  Indels    0;  Gaps    0;

Qy    209 cgacttctgccgctgctgcatcccctgcatcacccctgtttcagcgccagccaaggcag 268
          || | ||  |||   |  | ||  |  |   |  | | |     ||| | |    ||| |
Db    255 CGTCCTCGTCCGGGACCTCCTCGTCCTCGGGAAGCGCGGGCGTGGCGGCGGGGTCGGCGG 196

Qy    269 ccgcgcctcctgccgcggcgcgatcggctgagccgcacgtggggagcaagatcattgcta 328
          || | |  | ||||| ||| |  |||||   |||  || || || || | |   |  || |
Db    195 TCGGGGCCGCCGCCGCCGCGGGGGCGGCTCGTCCGGCCGCGGCGATCACGGCGAGGGCGA 136

Qy    329 atctagtgaatcagctggggattaatgtcacccaaaggagcgtcgtcagcactggagcgc 388
              |    |    | |   | |          | | | |  || | | ||
Db    135 GGAGGACGGGGAGGGGAAGAAAGGAGGGGGGGGTGCCGCGAGGCCGCATCCCGGGGCGCG 76

Qy    389 cggccacgacgaggtctacggcggtgacgtccacgactaccgccccgcagcgaac 443
          ||| | | ||| || | ||  ||  | |||| ||  |  ||  || || |||| |
Db     75 CGGGCGCTACGCGGGCGACTACGCGGGCGTCGACTCCGGCCCGCCGGCCGCGAGC 21


RESULT   8
AAA59242
ID   AAA59242 standard; cDNA; 1037 BP.
XX
AC   AAA59242;
XX
DT   07-NOV-2000  (first entry)
XX
DE   Exons E, C and A of the SCA8 repeat region.
XX
KW   Repeat region; spinocerebellar ataxia type 8; SCA8; long arm;
KW   chromosome 13; polymorphic CTA repeat; CTG repeat; ss.
XX
OS   Homo sapiens.
XX
PN   CA2283758-A1.
XX
PD   28-APR-2000.
XX
PF   27-SEP-1999;   99CA-2283758.
XX
PR   28-OCT-1998;   98US-0181585.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Benzow KA,  Moseley-alldredge ML,  Ranum LPW,  Koob MD;
XX
DR   WPI; 2000-491456/44.
XX
PT   Novel nucleic acid molecule containing repeat region of spinocerebellar
PT   ataxia type 8 coding sequence, useful as primer or probe for diagnosing
PT   or determining risk for developing spinocerebellar ataxia -
XX
PS   Claim 25; Fig 7C; 68pp; English.
XX
CC   The present sequence represents a nucleic acid molecule containing a
CC   repeat region of a spinocerebellar ataxia type 8 (SCA8) coding
CC   sequence which is located within the long arm of chromosome 13. The
CC   SCA8 coding sequence contains polymorphic CTA repeats and CTG repeats.
CC   Generally, the CTA repeat is unstable, and can vary between individuals
CC   in different families (the number of CTA repeats in the region does not
CC   vary between individuals of a family). Fragments of the nucleic acid
CC   molecule are useful as primers and probes for determining whether an
CC   individual has or is at risk of developing spinocerebellar ataxia 8.
XX
SQ   Sequence 1037 BP; 229 A; 261 C; 242 G; 305 T; 0 other;


  Query Match             3.9%;  Score 38;  DB 21;  Length 1037;
  Best Local Similarity   56.3%;  Pred. No. 0.67;
  Matches   71;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy    173 cgggcactgacgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatccc 232
          | |||  |    |  | | ||  | |  ||| | ||| ||| ||||  |||||||  |
Db    629 ctggctttactactactactactactactactactactactgctactgctgctgctg 688

Qy    233 ctgcatcacccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgat 292
          ||||  |  |  ||| | | ||  | ||    || || ||  || |||| || ||
Db    689 ctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 748

Qy    293 cggctg 298
          | ||||
Db    749 ctgctg 754


RESULT   9
AAA59240
ID   AAA59240 standard; DNA; 1159 BP.
XX
AC   AAA59240;
XX
DT   07-NOV-2000  (first entry)
XX
DE   An EcoRI fragment that includes SCA8 repeat region.
XX
KW   Repeat region; spinocerebellar ataxia type 8; SCA8; long arm;
KW   chromosome 13; polymorphic CTA repeat; CTG repeat; ss.
XX
OS   Homo sapiens.
XX
PN   CA2283758-A1.
XX
PD   28-APR-2000.
XX
PF   27-SEP-1999;   99CA-2283758.
XX
PR   28-OCT-1998;   98US-0181585.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Benzow KA,  Moseley-alldredge ML,  Ranum LPW,  Koob MD;
XX
DR   WPI; 2000-491456/44.
XX
PT   Novel nucleic acid molecule containing repeat region of spinocerebellar
PT   ataxia type 8 coding sequence, useful as primer or probe for diagnosing
PT   or determining risk for developing spinocerebellar ataxia -
XX
PS   Claim 26; Fig 7A; 68pp; English.
XX
CC   The present sequence represents a nucleic acid molecule containing a
CC   repeat region of a spinocerebellar ataxia type 8 (SCA8) coding
CC   sequence which is located within the long arm of chromosome 13. The
CC   SCA8 coding sequence contains polymorphic CTA repeats and CTG repeats.
CC   Generally, the CTA repeat is unstable, and can vary between individuals
CC   in different families (the number of CTA repeats in the region does not
CC   vary between individuals of a family). Fragments of the nucleic acid
CC   molecule are useful as primers and probes for determining whether an
CC   individual has or is at risk of developing spinocerebellar ataxia 8.
XX
SQ   Sequence 1159 BP; 263 A; 263 C; 196 G; 437 T; 0 other;


  Query Match             3.9%;  Score 38;  DB 21;  Length 1159;
  Best Local Similarity   56.3%;  Pred. No. 0.7;
  Matches   71;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;
```

```
Qy    173 cgggcactgacgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatccc 232
          | |||  |    |  | | ||  | |  ||| | ||| ||| ||||  |||||||  |
Db    441 ctggctttactactactactactactactactactactactgctactgctgctgctg 500

Qy    233 ctgcatcacccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgat 292
          ||||  |  |  ||| | | ||  | ||    || || ||  || |||| || ||
Db    501 ctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 560

Qy    293 cggctg 298
          | ||||
Db    561 ctgctg 566
```

RESULT 10
AAA59241
ID AAA59241 standard; cDNA; 1472 BP.
XX
AC AAA59241;
XX
DT 07-NOV-2000 (first entry)
XX
DE Exons D, C, B and A of the SCA8 repeat region.
XX
KW Repeat region; spinocerebellar ataxia type 8; SCA8; long arm;
KW chromosome 13; polymorphic CTA repeat; CTG repeat; ss.
XX
OS Homo sapiens.
XX
PN CA2283758-A1.
XX
PD 28-APR-2000.
XX
PF 27-SEP-1999; 99CA-2283758.
XX
PR 28-OCT-1998; 98US-0181585.
XX
PA (MINU) UNIV MINNESOTA.
XX
PI Benzow KA, Moseley-alldredge ML, Ranum LPW, Koob MD;
XX
DR WPI; 2000-491456/44.
XX
PT Novel nucleic acid molecule containing repeat region of spinocerebellar
PT ataxia type 8 coding sequence, useful as primer or probe for diagnosing
PT or determining risk for developing spinocerebellar ataxia -
XX
PS Claim 24; Fig 7B; 68pp; English.
XX
CC The present sequence represents a nucleic acid molecule containing a
CC repeat region of a spinocerebellar ataxia type 8 (SCA8) coding
CC sequence which is located within the long arm of chromosome 13. The
CC SCA8 coding sequence contains polymorphic CTA repeats and CTG repeats.
CC Generally, the CTA repeat is unstable, and can vary between individuals
CC in different families (the number of CTA repeats in the region does not
CC vary between individuals of a family). Fragments of the nucleic acid
CC molecule are useful as primers and probes for determining whether an
CC individual has or is at risk of developing spinocerebellar ataxia 8.
XX
SQ Sequence 1472 BP; 337 A; 368 C; 378 G; 388 T; 1 other;

```
  Query Match             3.9%;  Score 38;  DB 21;  Length 1472;
  Best Local Similarity   56.3%;  Pred. No. 0.76;
  Matches   71;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy    173 cgggcactgacgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatccc 232
          | |||  |    |  |·| ||  | |  ||| | ||| ||| ||||  |||||||  |
Db   1064 ctggctttactactactactactactactactactactactgctactgctgctgctg 1123

Qy    233 ctgcatcacccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgat 292
          ||||  |  |  ||| | | ||  | ||    || || ||  || |||| || ||
Db   1124 ctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1183

Qy    293 cggctg 298
          | ||||
Db   1184 ctgctg 1189
```

RESULT 11
AAZ19383
ID AAZ19383 standard; cDNA; 371 BP.
XX
AC AAZ19383;
XX
DT 05-NOV-1999 (first entry)
XX
DE M. tuberculosis antigen 5' LSER-10 cDNA sequence.
XX
KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
KW immunotherapy; diagnosis; immunisation; vaccine; infection;
KW immune response; skin test; ss.
XX
OS Mycobacterium tuberculosis.
XX
PN WO9942076-A2.
XX
PD 26-AUG-1999.
XX
PF 17-FEB-1999; 99WO-US03268.
XX
PR 05-MAY-1998; 98US-0072967.
PR 18-FEB-1998; 98US-0025197.
XX
PA (CORI-) CORIXA CORP.
XX
PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
DR WPI; 1999-527409/44.
DR P-PSDB; AAY39190.
XX
PT New antigens from Mycobacterium tuberculosis useful in diagnostic
PT skin tests and protective or therapeutic vaccines or compositions
XX
PS Claim 4; Page 221; 299pp; English.
XX
CC The present invention describes polypeptides comprising an immunogenic
CC part of a Mycobacterium tuberculosis antigen (Ag). Also described
CC are vaccines and fusion protein containing M. tuberculosis Ag's.
CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
CC other polypeptides fragments, can be used in pharmaceutical compositions
CC or vaccines to generate a protective or therapeutic immune response to
CC M. tuberculosis and as reagents in skin tests for diagnosis of
CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
CC by, T, B or natural killer cells and/or macrophages in
CC tuberculosis-immune subjects. AAZ19249 to AAZ19460 and AAY39083 to
CC AAY39225 are used in the exemplification of the present invention.
XX
SQ Sequence 371 BP; 50 A; 128 C; 163 G; 30 T; 0 other;

```
  Query Match             3.9%;  Score 37.8;  DB 20;  Length 371;
  Best Local Similarity   47.3%;  Pred. No. 0.52;
  Matches  114;  Conservative    0;  Mismatches  127;  Indels    0;  Gaps    0;

Qy    373 gtcagcactggagcgccggccacgacgaggtctacggcggtgacgtccacgactaccgcc 432
          || | ||| |  |  | | | |  | ||  || |  |||| |   ||  |  || |  | ||
Db      1 gtgaccacggtgggcctgccaccaacccgggcagcggcagccgcggcggcgccggcggct 60

Qy    433 ccgcagcgaacaagtccatacgggcacaatggccgacctgtgacggctggattggtggca 492
          |||  |  |||     |    |||   |     ||   | |||| |   || | |
Db     61 ccggcggcaacggtggcgccgggggtaacgccaccggctcaggcggcaagggcggccgcg 120

Qy    493 gctaatagtggtgccagcgcggcctcgtctcccacagccgcggcgaaaccaacaggagaa 552
          |     |  |  |  |  |  || ||| | |   |||| ||  ||   ||| | ||   |
```

```
Db    121 gtggcaatggcggtgatgggagcttcggcgctaccagcggccccgcctccatcggggtca 180

Qy    553 gaaaaggcctccgcggcatgtgaaacgagctccgtggcgataaatgcgacacgcccggcg 612
              | || | ||||||   |   | ||  | | | || |    || || |   ||||
Db    181 cgggcgcccccggcggcaacggcggcaagggcggcgccggtggcagcaaccccaacggct 240

Qy    613 c 613
          |
Db    241 c 241


RESULT  12
AAZ19171
ID   AAZ19171 standard; cDNA; 371 BP.
XX
AC   AAZ19171;
XX
DT   05-NOV-1999  (first entry)
XX
DE   M. tuberculosis recombinant antigen cDNA encoding 5' LSER-10.
XX
KW   Antigen; diagnosis; detection; infection; antibody; immunisation;
KW   vaccine; immunity; ss.
XX
OS   Mycobacterium tuberculosis.
XX
PN   WO9942118-A2.
XX
PD   26-AUG-1999.
XX
PF   17-FEB-1999;   99WO-US03265.
XX
PR   05-MAY-1998;   98US-0072596.
PR   18-FEB-1998;   98US-0024753.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Campos-Neto A,  Dillon DC,  Hendrickson RC,  Houghton R;
PI   Lodes MJ,  Reed SG,  Skeiky YAW,  Twardzik DR,  Vedvick TS;
XX
DR   WPI; 1999-527416/44.
DR   P-PSDB; AAY39047.
XX
PT   New polypeptide comprising antigenic portions of M. tuberculosis
XX
PS   Claim 4; Page 266; 323pp; English.
XX
CC   This invention describes novel recombinant antigens and their encoding
CC   nucleic acids derived from Mycobacterium tuberculosis. The novel
CC   polypeptides are useful for detecting M. tuberculosis infection in a
CC   biological sample by detecting antibodies which bind with the
CC   polypeptides, and are useful as vaccines for immunizing against
CC   M. tuberculosis infection. The new detection methods are needed as
CC   current vaccination strategies do not provide 100% immunity.
XX
SQ   Sequence 371 BP; 50 A; 128 C; 163 G; 30 T; 0 other;

  Query Match             3.9%;  Score 37.8;  DB 20;  Length 371;
  Best Local Similarity   47.3%;  Pred. No. 0.52;
  Matches  114;  Conservative    0;  Mismatches  127;  Indels    0;  Gaps    0;

Qy    373 gtcagcactggagcgccggccacgacgaggtctacggcggtgacgtccacgactaccgcc 432
          || | ||| |  |  | | |  | ||  || |  |||| |   || |  || |  | ||
Db      1 gtgaccacggtgggcctgccaccaacccgggcagcggcagccgcggcggcgccggcggct 60

Qy    433 ccgcagcgaacaagtccatacgggcacaatggccgacctgtgacggctggattggtggca 492
          |||  |  |||     |    |||   |     |   ||  | ||||  |   || | |
Db     61 ccggcggcaacggtggcgccgggggtaacgccaccggctcaggcggcaagggcggcgccg 120

Qy    493 gctaatagtggtgccagcgcggcctcgtctcccacagccgcggcgaaaccaacaggagaa 552
          |     |  |  |  |  |  || ||| | |   |||| ||  ||   ||| | ||   |
Db    121 gtggcaatggcggtgatgggagcttcggcgctaccagcggccccgcctccatcggggtca 180

Qy    553 gaaaaggcctccgcggcatgtgaaacgagctccgtggcgataaatgcgacacgcccggcg 612
              | || | ||||||   |   | ||  | | | || |    || || |   ||||
Db    181 cgggcgcccccggcggcaacggcggcaagggcggcgccggtggcagcaaccccaacggct 240

Qy    613 c 613
          |
Db    241 c 241


RESULT  13
AAZ32263
ID   AAZ32263 standard; cDNA; 2115 BP.
XX
AC   AAZ32263;
XX
DT   19-JAN-2000  (first entry)
XX
DE   Human dunce-like phosphodiesterase pPDE39 encoding cDNA.
XX
KW   Phosphodiesterase; dunce-like phosphodiesterase; PDE; DPD; cAMP;
KW   RAS-related protein; immunoreactive; detection; genetic defect;
KW   bronchodilation; increased myocardial contractility;
KW   anti-inflammation; ss.
XX
OS   Homo sapiens.
XX
PN   US5977305-A.
XX
PD   02-NOV-1999.
XX
PF   07-JUN-1995;   95US-0474379.
XX
PR   01-MAR-1994;   94US-0206188.
PR   20-APR-1990;   90US-0511715.
PR   19-APR-1991;   91US-0688352.
XX
PA   (COLD-) COLD SPRING HARBOR LAB.
XX
PI   Colicelli JJ,  Wigler MH;
XX
DR   WPI; 1999-619709/53.
DR   P-PSDB; AAY49826.
XX
PT   New isolated RAS-related polypeptides and mammalian cyclic nucleotide
PT   phosphodiesterases, used for screening for agents which can modify
PT   complement or suppress genetic defects
XX
PS   Example 6; Column 165-170; 145pp; English.
XX
CC   The present invention describes new isolated RAS-related polypeptides
CC   and mammalian cyclic nucleotide phosphodiesterases (PDEs). RAS-related
CC   polypeptides are capable of complementing a defective RAS function in
CC   yeast. The products can be used for screening for agents which can
CC   modify, complement or suppress a genetic defect in a biochemical
CC   pathway in which cAMP participates, or in a biochemical pathway which
CC   is controlled, directly or indirectly, by a RAS protein and other
CC   proteins affecting cell growth and maintenance. Developing agents that
CC   will selectively act upon PDEs is directed toward reproducing the
CC   desirable effects of cyclic nucleotides, e.g. bronchodilation,
CC   increased myocardial contractility, anti-inflammation, yet without
CC   causing the undesirable effects, e.g. increased heart rate or enhanced
CC   lipolysis. The products can also be used for therapeutic, diagnostic
CC   and prognostic uses. AAZ32229 to AAZ32285, and AAY49803 to AAY49830,
CC   represent sequences used in the exemplification of the present
CC   invention.
XX
SQ   Sequence 2115 BP; 536 A; 605 C; 501 G; 473 T; 0 other;

  Query Match             3.9%;  Score 37.6;  DB 20;  Length 2115;
  Best Local Similarity   63.0%;  Pred. No. 1.1;
  Matches   58;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;
```

```
Qy    212 cttctgccgctgctgcatcccctgcatcacccctgtttcagcgccagccaaggcagccg 271
          || |  |||| || ||  |||  ||  || |||| |   ||| | |||||   || ||||
Db    577 ctgcccccgccgccgccaccctcgccccagccccagccccagtgtcagccgccgccgccg 636

Qy    272 cgcctcctgccgcggcgcgatcggctgagccg 303
          | || ||||||||| || |   | || |  | |
Db    637 ccccccctgccgccgcccccgccgccgcccgg 668


RESULT 14
AAA88198
ID   AAA88198 standard; cDNA; 2115 BP.
XX
AC   AAA88198;
XX
DT   14-DEC-2000  (first entry)
XX
DE   pPDE39 human dunce-like PDE nucleotide sequence SEQ ID NO:60.
XX
KW   Detection; mammalian gene; yeast; microorganism; identification;
KW   phenotype; characteristic; dunce-like phosphodiesterase; PDE; RAS;
KW   RAS-related protein; genetic defect; hybridisation; probe; ss.
XX
OS   Homo sapiens.
OS   Saccharomyces cerevisiae.
XX
PN   US6100025-A.
XX
PD   08-AUG-2000.
XX
PF   01-MAR-1994;   94US-0206188.
XX
PR   20-APR-1990;   90US-0511715.
PR   19-APR-1991;   91US-0688352.
XX
PA   (COLD-) COLD SPRING HARBOR LAB.
XX
PI   Colicelli JJ,  Wigler MH;
XX
DR   WPI; 2000-531664/48.
DR   P-PSDB; AAB20637.
XX
PT   Novel isolated DNA encoding a mammalian cyclic nucleotide
PT   phosphodiesterase is present in plasmids pPDE46, pPDE43 or pPDE339 and
PT   is used to modify a genetic defect in a biochemical pathway in which
PT   cAMP participates  -
XX
PS   Claim 2; Column 171-176; 145pp; English.
XX
CC   The present invention describes a purified and isolated DNA (I) which
CC   encodes a mammalian cyclic nucleotide phosphodiesterase and is an insert
CC   present in the plasmids pPDE46 (ATCC 69552), pPDE43 (ATCC 69551) or
CC   pPDE339 (ATCC 69550). The DNA molecules are used to modify, complement
CC   or suppress a genetic defect in a biochemical pathway in which cAMP
CC   participates and are also used as hybridisation probes. The present
CC   invention also describes methods for detecting mammalian genes encoding
CC   proteins which can function in microorganisms, particularly yeast, to
CC   modify, complement, or suppress a genetic defect associated with an
CC   identifiable phenotypic alteration or characteristic in the
CC   microorganism. AAA88162 to AAA88218 and AAB29614 to AAB20640 represent
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence 2115 BP; 536 A; 605 C; 501 G; 473 T; 0 other;


  Query Match             3.9%;  Score 37.6;  DB 21;  Length 2115;
  Best Local Similarity   63.0%;  Pred. No. 1.1;
  Matches   58;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy    212 cttctgccgctgctgcatcccctgcatcacccctgtttcagcgccagccaaggcagccg 271
          || |  |||| || ||  |||  ||  || |||| |   ||| | |||||   || ||||
Db    577 ctgcccccgccgccgccaccctcgccccagccccagccccagtgtcagccgccgccgccg 636

Qy    272 cgcctcctgccgcggcgcgatcggctgagccg 303
          | || ||||||||| || |   | || |  | |
Db    637 ccccccctgccgccgcccccgccgccgcccgg 668


RESULT 15
AAA81468
ID   AAA81468 standard; DNA; 72750 BP.
XX
AC   AAA81468;
XX
DT   04-DEC-2000  (first entry)
XX
DE   N. meningitidis partial DNA sequence gnm_16 SEQ ID NO:16.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; genome; immunogenic;
KW   antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW   Meningococcus B; MenB; ds.
XX
OS   Neisseria meningitidis.
XX
PN   WO200022430-A2.
XX
PD   20-APR-2000.
XX
PF   08-OCT-1999;   99WO-US23573.
XX
PR   09-OCT-1998;   98US-0103794.
PR   30-APR-1999;   99US-0132068.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Frazer CM,  Hickey E,  Peterson J,  Tettelin H,  Venter JC;
PI   Masignani V,  Galeotti C,  Mora M,  Ratti G,  Scarselli M,  Scarlato V;
PI   Rappuoli R,  Pizza M;
XX
DR   WPI; 2000-318079/27.
XX
PT   Isolated nucleotide sequences of Neisseria meningitidis which can be
PT   used in the diagnosis and treatment of N. meningitidis infection and
PT   other Neisserial infections, for example, N.gonorrhoea -
XX
PS   Claim 7; Page 406-428; 1760pp; English.
XX
CC   The present invention describes methods of obtaining immunogenic
CC   proteins from Neisseria genomic sequences. AAA81453 to AAA82414
CC   represent specifically claimed Neisseria meningitidis genomic DNA
CC   sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
CC   Neisseria DNA sequences and their corresponding proteins; AAA81254 to
CC   AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
CC   isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
CC   AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
CC   sequences, which are all used in the exemplification of the present
CC   invention. The nucleic acid sequences, protein sequences, and antibodies
CC   against them, can be used in the manufacture of a composition. The
CC   composition can be used as a medicament (or in the manufacture of a
CC   medicament) for treating, preventing or diagnosing infection due to
CC   Neisserial bacteria. For example, some of the identified proteins could
CC   be components of vaccines against Meningococcus B; against all serotypes;
CC   and/or against all pathogenic Neissariae. Identification of sequences
CC   from the bacterium will also facilitate production of biological probes,
CC   particularly organism-specific probes. Attempts to make efficacious
CC   Meningococcus B vaccines have failed mainly due to antigen tolerance.
CC   Multivalent vaccines have also been tried but none have successfully
CC   overcome antigenic variability. The provision of further, complete
CC   sequences may provide an opportunity to identify secreted or surface
CC   exposed proteins that may be presumed targets for the immune system and
CC   which are not antigenically variable or at least more conserved than
CC   other more variable regions.
XX
SQ   Sequence 72750 BP; 17518 A; 19945 C; 18810 G; 16477 T; 0 other;
```

Query Match 3.9%; Score 37.6; DB 21; Length 72750;
Best Local Similarity 55.3%; Pred. No. 4.1;
Matches 73; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

```
Qy     395 cgacgaggtctacggcggtgacgtccacgactaccgccccgcagcgaacaagtccatacg 454
           ||| | ||   |   |  | ||| |  ||||| ||| | |   |  || | ||   | |
Db    9250 cgaagtggatgaactcactcacggcggcgacttccggcacttcggcaaaacgttttttca 9309

Qy     455 ggcacaatggccgacctgtgacggctggattggtggcagctaatagtggtgccagcgcgg 514
           | ||  ||       || ||||  |  |||||||||  |||   |  | |   |  ||||
Db    9310 gccaatattcgatggctttgacatcgtgattggtggtggcttcgatggctttgacggcgg 9369

Qy     515 cctcgtctccca 526
           || ||||| |||
Db    9370 ccgcgtcttcca 9381
```

Search completed: September 21, 2001, 02:42:32
Job time: 11448 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: September 20, 2001, 23:27:46 ; Search time 4309.39 Seconds
(without alignments)
2140.903 Million cell updates/sec

Title: US-09-138-735-1_COPY_1232_2207
Perfect score: 976
Sequence: 1 caggtacagcgtaacggctt.....tgtgaacgaggccattatgt 976

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est75:*
137: gb_est76:*
138: gb_est77:*
139: gb_est78:*
140: gb_est79:*
141: gb_est80:*
142: gb_est81:*
143: gb_est82:*
144: gb_est83:*
145: gb_est84:*
146: gb_est85:*
147: gb_est86:*
148: gb_est87:*
149: gb_est88:*
150: gb_est89:*
151: gb_est90:*
152: gb_est99:*
153: gb_est100:*
154: gb_est101:*
155: gb_est102:*
156: gb_est103:*
157: gb_est104:*
158: gb_est105:*
159: gb_est106:*
160: gb_est67:*
161: gb_est68:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_esthum29:*
177: em_esthum30:*
178: em_esthum31:*
179: em_esthum32:*
180: em_esthum33:*
181: em_estom3:*
182: em_estpl11:*
183: em_estro21:*
184: em_estro22:*
185: em_estro23:*
186: em_htc:*
187: gb_est107:*
188: gb_est108:*
189: gb_est109:*
190: gb_est110:*
191: gb_est111:*
192: gb_htc:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_other:*
207: em_gss_pln1:*
208: em_gss_pln2:*
209: em_gss_pro:*
210: em_gss_rod1:*
211: em_gss_rod2:*
212: em_gss_rod3:*
213: em_gss_rod4:*
214: em_gss_rod5:*
215: em_gss_vrt1:*
216: em_gss_vrt2:*
217: em_gss_vrt3:*
218: gb_gss1:*
219: gb_gss2:*
220: gb_gss3:*
221: gb_gss4:*
222: gb_gss5:*
223: gb_gss6:*
224: gb_gss7:*
225: gb_gss8:*
226: gb_gss9:*
227: gb_gss10:*
228: gb_gss11:*
229: gb_gss12:*
230: gb_gss13:*
231: gb_gss14:*
232: gb_gss15:*
233: gb_gss16:*
234: gb_gss17:*
235: gb_gss18:*
236: gb_gss19:*
237: gb_gss20:*
238: gb_gss21:*
239: gb_gss22:*
240: gb_gss23:*
241: gb_gss24:*
242: gb_gss25:*
243: gb_gss26:*
244: gb_gss27:*
245: gb_gss28:*
246: gb_gss29:*
247: gb_gss30:*
248: gb_gss31:*
249: gb_gss32:*
250: gb_gss33:*
251: gb_gss34:*
252: em_gss_inv4:*
253: em_gss_rod6:*
254: em_gss_rod7:*
255: em_gss_rod8:*
256: gb_gss35:*
257: gb_gss36:*
258: gb_gss37:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	52.6	5.4	925	219	CNS0091P	AL053013 Drosophil
2	45.4	4.7	516	138	BE675897	BE675897 7f18b01.x
3	44	4.5	466	122	AW922844	AW922844 DG1_46_C0
4	43.8	4.5	1036	219	CNS010BS	AL098770 Drosophil
5	43.4	4.4	558	119	AW677864	AW677864 WS1_11_G0
6	43.4	4.4	1065	219	CNS0108S	AL098662 Drosophil
7	43.2	4.4	1101	219	CNS017SY	AL108460 Drosophil
8	43	4.4	577	137	BE587138	BE587138 WHE0517_A
9	42.8	4.4	925	219	CNS0091P	AL053013 Drosophil
10	42.6	4.4	787	219	CNS010B7	AL098749 Drosophil
11	42.4	4.3	504	138	BE675949	BE675949 7f19a02.x
12	42.4	4.3	577	167	BE401994	BE401994 CSB003C12
c 13	42.2	4.3	390	145	BF152676	BF152676 uz94e04.y
14	41.8	4.3	391	137	BE595680	BE595680 PI1_55_C0
15	41.8	4.3	488	120	AW747489	AW747489 WS1_68_D0
16	41.8	4.3	554	119	AW677991	AW677991 WS1_12_H0
17	41.6	4.3	500	30	AV437560	AV437560 AV437560
18	41.4	4.2	487	154	BG466260	BG466260 EtESTed50
19	41.4	4.2	559	247	AZ661219	AZ661219 1M0539M10
c 20	41.2	4.2	447	12	AA858891	AA858891 UI-R-A0-b
21	40.2	4.1	332	166	BE363235	BE363235 WS1_61_G0
22	40.2	4.1	339	122	AW925156	AW925156 WS1_76_B0
23	40.2	4.1	352	137	BE595557	BE595557 PI1_53_B0
24	40.2	4.1	353	166	BE363396	BE363396 WS1_62_E0
25	40.2	4.1	410	137	BE592380	BE592380 WS1_94_A0
26	40.2	4.1	412	166	BE356035	BE356035 DG1_121_D
27	40.2	4.1	418	166	BE363254	BE363254 WS1_61_F0
28	40.2	4.1	444	166	BE362725	BE362725 DG1_88_B0
29	40.2	4.1	447	122	AW922332	AW922332 DG1_17_C0
30	40.2	4.1	448	155	BG557134	BG557134 EM1_43_A0
31	40.2	4.1	459	120	AW745877	AW745877 WS1_38_A0
32	40.2	4.1	460	166	BE363431	BE363431 WS1_62_H0
33	40.2	4.1	462	137	BE592941	BE592941 WS1_92_C0
34	40.2	4.1	480	137	BE592157	BE592157 WS1_89_B0
35	40.2	4.1	489	137	BE594042	BE594042 WS1_101_E
36	40.2	4.1	489	166	BE363359	BE363359 WS1_62_A0
37	40.2	4.1	490	154	BG464372	BG464372 EM1_69_A0
38	40.2	4.1	493	152	BG356444	BG356444 EM1_22_F0
39	40.2	4.1	497	119	AW677984	AW677984 WS1_12_H1
40	40.2	4.1	498	155	BG556692	BG556692 EM1_40_D0
41	40.2	4.1	499	166	BE363274	BE363274 WS1_61_E0
42	40.2	4.1	500	141	BE917797	BE917797 OV1_7_C02
43	40.2	4.1	502	137	BE593867	BE593867 WS1_103_D
44	40.2	4.1	502	166	BE356939	BE356939 DG1_145_H
45	40.2	4.1	503	137	BE593309	BE593309 WS1_100_B

ALIGNMENTS

```
RESULT   1
CNS0091P/c
LOCUS       CNS0091P      925 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL053013
VERSION     AL053013.1  GI:4934461
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 925)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .925
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR19D16"
                     /note="end : TET3"
BASE COUNT      120 a      61 c      61 g     172 t     511 others
ORIGIN
```

Query Match 5.4%; Score 52.6; DB 219; Length 925;
Best Local Similarity 14.0%; Pred. No. 0.0018;
Matches 55; Conservative 172; Mismatches 166; Indels 0; Gaps 0;

```
Qy    221 ctgctgcatcccctgcatcacccctgtttcagcgccagccaaggcagccgcgcctcctg 280
          | :: :|  :|::::|  : :: ::    :| :|::: :::  ::: ::::::: :: :
Db    925 CSBSCSCSCSCSSBSCSSSSMSTSSSNSSBCSSCSSSBSSSSTSSMSSSSBSSSSGSSSS 866

Qy    281 ccgcggcgcgatcggctgagccgcacgtggggagcaagatcattgctaatctagtgaatc 340
          ::| ::  ::   ::: |   ||:|:    :: ::  |  :|  :  :|   ||::   :
Db    865 SSGTSSACVKCNASSSCGCCGCGMABCCMCSSSSSCCGSASARGVKVRASGGAGKRGGGS 806

Qy    341 agctggggattaatgtcacccaaaggagcgtcgtcagcactggagcgccggccacgacga 400
            |   : :   | :: : :|:|   :|::: :  : :::: |||| :  :::  :: ::
Db    805 GGASASHSSSSACBSSSSSCSASCWSASSSSSASSSRSRSGGGAGGSGASSSRSSSSSSS 746

Qy    401 ggtctacggcggtgacgtccacgactaccgccccgcagcgaacaagtccatacgggcaca 460
            :  :::: :: : :  :::: : |:: :|:::: ::  |:  : ::   | : : |
Db    745 ASAGSVVSSASSSSSSCSSSVSCSSVASSMSCSSBSSSSASASSSSSSSSASCASCSCCC 686

Qy    461 atggccgacctgtgacggctggattggtggcagctaatagtggtgccagcgcggcctcgt 520
          :|:   : : :| ::: ::    : ::: :  |  :  : ::: |:::: : ||:
Db    685 TSWSCSCSTSASMSAARSSSSSSSSCSSSSMSASSSASSSSASSSSSSGSSSSSGSACGB 626

Qy    521 ctcccacagccgcggcgaaaccaacaggagaagaaaaggcctccgcggcatgtgaaacga 580
          : ::    |::: ::     ::    :: |  |   :||:  ::: :|:  | | : :
Db    625 SMSSGGGSGSVSASSGMSSSVSSSGGRSSGSGGGGGVGGSGSSSSGSGSSGGSGSVCSCS 566

Qy    581 gctccgtggcgataaatgcgacacgcccggcgc 613
          :  :|: : ::| ||| :|:|  ||::||:::
Db    565 SGCMCRCSCSSAAAAAASCVAASCGMMCGKSKS 533
```

```
RESULT   2
BE675897
LOCUS       BE675897      516 bp    mRNA            EST       08-SEP-2000
DEFINITION  7f18b01.x1 NCI_CGAP_CLL1 Homo sapiens cDNA clone IMAGE:3294985 3',
            mRNA sequence.
ACCESSION   BE675897
VERSION     BE675897.1  GI:10036438
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 516)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever,
            M.D., Louis M. Staudt, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            Seq primer: -40UP from Gibco
            High quality sequence stop: 474.
FEATURES             Location/Qualifiers
     source          1. .516
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3294985"
                     /clone_lib="NCI_CGAP_CLL1"
                     /tissue_type="B-cell, chronic lymphotic leukemia"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATTGCTTTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library is normalized, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT      166 a     95 c    121 g    134 t
ORIGIN
```

```
  Query Match             4.7%;  Score 45.4;  DB 138;  Length 516;
  Best Local Similarity   49.0%;  Pred. No. 0.14;
  Matches  121;  Conservative    0;  Mismatches  126;  Indels    0;  Gaps    0;

Qy    536 cgaaaccaacaggagaagaaaaggcctccgcggcatgtgaaacgagctccgtggcgataa 595
          | |||||||    | ||  |||    ||  ||  |    || | |||   ||      ||
Db    235 CAAAACCAAACCAAAAAACAAATCGGTCAACGTTACCAAAACCAAGCAGAGTCTAACAAA 294

Qy    596 atgcgacacgcccggcgcttcacaacgcctctctcccgcaggcgccaacggatggcgttt 655
           ||  |||  || |  |   ||| | | ||   | || | || ||   | |||||  ||
Db    295 CTGTCACAGCCCAGAGGAGGCACGAGGACTGAATGCCACGGGGGCTCCCAGATGGGATTC 354

Qy    656 tggcggcagcagtataccagtcggagggcgaggttcatcagtcgctggagcggctggagt 715
          |||  | || |    |   |  ||  |   |   ||| |   | | || | |  ||
Db    355 TGGGAGAAGAAAAGGATGTGAGTGAAAACTGCGGAAATCTGAACCAAGTGCAGGCGTGGT 414

Qy    716 ccgtcataaccaacacgtctcgggttctgaagttgctccctgacaccattcgaagagacc 775
            |||||||  |    | | || || | || |    |    |  |||| | ||||||
Db    415 TAATCATAACACATCATTATTGGTTTGTTAACTGTGACAAATATGCCATACTAAGAGAAG 474

Qy    776 atgaaca 782
          | |  ||
Db    475 ACGGTCA 481
```

```
RESULT   3
AW922844
LOCUS       AW922844      466 bp    mRNA            EST       19-JUL-2000
DEFINITION  DG1_46_C04.g1_A002 Dark Grown 1 (DG1) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   AW922844
VERSION     AW922844.1  GI:8088669
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 466)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: dark-grown seedlings
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: PolyTMix
            High quality sequence start: 135
            High quality sequence stop: 466
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .466
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Dark Grown 1 (DG1)"
                     /note="Organ: 5-day-old dark-grown seedlings; Vector:
                     Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was
                     made from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT      67 a    121 c    182 g     95 t      1 others
ORIGIN
```

```
  Query Match             4.5%;  Score 44;  DB 122;  Length 466;
  Best Local Similarity   49.8%;  Pred. No. 0.32;
  Matches  110;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy    210 gacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaaggcagc 269
          | || | |  || |  ||  |  || |  | ||||| ||    | | | || | ||  |
Db     30 GGCTACGGAGGCGGTGGCGGCGGCTACGGCGCCCCCCGTGAGGGTGNCGGCTACGGAGGA 89

Qy    270 cgcgcctcctgccgcggcgcgatcggctgagccgcacgtggggagcaagatcattgctaa 329
            ||| |  | ||  ||||| |  | ||   |  | | | ||  |   ||    | ||||
Db     90 GGCGGCGGCGGCTACGGCGGGCGCCGCGAGGGAGGAGGCGGCTACGGAGGCGGTGGCTAC 149

Qy    330 tctagtgaatcagctggggattaatgtcacccaaaggagcgtcgtcagcactggagcgcc 389
          | | |      |  ||  |     ||| ||  |||  |  |     |  | ||| |
Db    150 GGTGGCGGCGGCGGCGGCTACGGCGGTCCCCGTGAGGGCGGCGGCGGCTATGGCGGCGGC 209

Qy    390 ggccacgacgaggtctacggcggtgacgtccacgactaccg 430
          |||  | |||  | | || ||||||||  |  ||  | || |
Db    210 GGCGGCTACGGTGGCAACCGCGGTGACTCCGGCGGCAACTG 250
```

```
RESULT   4
CNS010BS
LOCUS       CNS010BS     1036 bp    DNA             GSS       26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN03H11 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL098770
VERSION     AL098770.1  GI:5610381
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Plasmid Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1036)
  AUTHORS   Genoscope.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1036
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN03H11"
                     /note="end : SP6"
BASE COUNT      240 a    127 c    155 g    234 t    280 others
ORIGIN


  Query Match             4.5%;  Score 43.8;  DB 219;  Length 1036;
  Best Local Similarity   23.6%;  Pred. No. 0.4;
  Matches   92;  Conservative  113;  Mismatches  185;  Indels    0;  Gaps    0;

Qy      94 cggtcctcccgcagggaagaacagcagccaggccaaaaaacatcggtagtggcgacggcg 153
           |:  || :: ::::|  |:|| |  ::  : |: |||||| | | | |: :::: : :||
Db     637 CKACCCSSSAVSMSGASARAAAAAAVSAAVAGMGAAAAAAGASCYGGASGRRSSSSCRCG 696

Qy     154 aaaccggggtgtgtgtcctcgggcactgacgcggcgagtagcagtcataccaatacgact 213
             :|:  :       :    |       || :  :   : ||||:    |:: |  : |:
Db     697 CSVCSYCKCGCGCCCYGAAAGAAAGACGAGSAAVGCGSAAGCASSAGCAMMGASCSAASA 756

Qy     214 tctgccgctgctgcatcccctgcatcacccctgtttcagcgccagccaaggcagccgcg 273
               |   :| :|: | :|     | : : : |   : :|:::  : ||| :|: :|
Db     757 RAVCGCCGCSCGSCVSCASCSCGCSCGSGMGSCGAMASSSCSSSCASWAAGASASGSGGC 816

Qy     274 cctcctgccgcggcgcgatcggctgagccgcacgtggggagcaagatcattgctaatcta 333
           :| |  ::|:::|:       :: ||||: | :  :| | :: :|  :|  |:  |
Db     817 SCSCGASSCVSSGSCGCGMGCVSCGAGCSCCSSCGSGCGSSSGMGCSSASGGSGSASASS 876

Qy     334 gtgaatcagctggggattaatgtcacccaaaggagcgtcgtcagcactggagcgccggcc 393
           : :     :: ::|    :: : : ::    ::|  |  :  |||:  |: ||||   :|
Db     877 SGSTSCGSSSGSSGCGASRMASGSGSSGGSGSVAARGCGSARAGCRRAGSSGCGCACNSC 936

Qy     394 acgacgaggtctacggcggtgacgtccacgactaccgccccgcagcgaacaagtccatac 453
           |:  |: |    |::::|   |     | :|:   :::|: :  ::|: :    : : |
Db     937 ASMGCKCGMAAAASVSSGMGCAAAAAAAARASGCGSSSCSGSGGSVGVSBSTCGSARVAR 996

Qy     454 gggcacaatggccgacctgtgacggctgga 483
           |:|  :|  |:::| :: : : :|:: |||
Db     997 GSGGCSAGCGSSSGGSSVSSVGSGSSGGGA 1026


RESULT   5
AW677864
LOCUS       AW677864      558 bp    mRNA            EST       19-JUL-2000
DEFINITION  WS1_11_G05.g1_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA,
            mRNA sequence.
ACCESSION   AW677864
VERSION     AW677864.1  GI:7551577
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 558)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: water-stressed plants
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: T7
            High quality sequence start: 146
            High quality sequence stop: 558
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .558
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Water-stressed 1 (WS1)"
                     /note="Organ: Mix of 5-week old plants on days 7 & 8 after
                     water was withheld; Vector: Lambda Zap; Site_1: XhoI;
                     Site_2: EcoRI; The library was made from poly-A RNA in the
                     cloning vector lambda ZAP II.  Clones to be sequenced were
                     prepared by mass excision."
BASE COUNT       88 a    175 c    210 g     85 t
ORIGIN


  Query Match             4.4%;  Score 43.4;  DB 119;  Length 558;
  Best Local Similarity   49.8%;  Pred. No. 0.47;
  Matches  110;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy     210 gacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaaggcagc 269
           | || | |  || |  ||  |  || |  | ||||| ||    | | | || | ||  |
Db     255 GGCTACGGAGGCGGTGGCGGCGGCTACGGCGCCCCCCGTGAGGGTGGCGGCTACGGAGGA 314

Qy     270 cgcgcctcctgccgcggcgcgatcggctgagccgcacgtggggagcaagatcattgctaa 329
           |||  |  | ||  ||||| |  | ||   |  | |   ||  |  |||    | ||||
Db     315 GGCGGCGGCGGCTACGGCGGGCGCCGCGAGGGAGGAGCCGGCTACGAAGGCGGTGGCTAC 374

Qy     330 tctagtgaatcagctggggattaatgtcacccaaaggagcgtcgtcagcactggagcgcc 389
           | | |      |  ||  |     ||| ||   |||   |  |     |  |  ||| |
Db     375 GGTGGCGGCGGCGGCGGCTACGGCGGTCGCCGTGAGGGCGGCGGCGGCTATGGCGGCGGC 434

Qy     390 ggccacgacgaggtctacggcggtgacgtccacgactaccg 430
           |||  | |||  | | || ||||||||  |  || | || |
Db     435 GGCGGCTACGGTGGCAACCGCGGTGACTCCGGCGGCAACTG 475


RESULT   6
CNS0108S
LOCUS       CNS0108S      1065 bp    DNA             GSS       26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN03J03 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL098662
VERSION     AL098662.1  GI:5610273
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Plasmid Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1065)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
```

```
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1065
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN03J03"
                     /note="end : T7"
BASE COUNT      121 a    225 c    252 g    313 t    154 others
ORIGIN


  Query Match             4.4%;  Score 43.4;  DB 219;  Length 1065;
  Best Local Similarity   27.4%;  Pred. No. 0.51;
  Matches   51;  Conservative   62;  Mismatches   73;  Indels    0;  Gaps    0;

Qy     141 agtggcgacggcgaaaccggggtgtgtgtcctcgggcactgacgcggcgagtagcagtca 200
           :: :|:: ::::|   :  :  |:||:|    ::| |  :| ::::: : :: |: |:
Db     755 RVGSGSSGSSSSGSGSSGCSTCTSTGKGGGGSSSGCCCGKGSSSSKBGSSBKSGSGGKGT 814

Qy     201 taccaatacgacttctgccgctgctgcatcccctgcatcacccctgtttcagcgccagc 260
           |||  |  |  |:| :  :    || :|||| |  | ||||      |: | || :|
Db     815 CACCTTTBKGTKTWCGKTTCSCCGGGCYYCCCCCCCCCCTCCCCGSCGNCCMCCCCCCSC 874

Qy     261 caaggcagccgcgcctcctgccgcggcgcgatcggctgagccgcacgtggggagcaagat 320
           |: | |  :::: |: |: :::::::::  ||: | : |::|: :| |::: ||  |::
Db     875 CMTGCCYTSSSSCCSSCSSSSSSSSSSSSCCTCKCCSKTGSSGSSSGGGSSSCGCTKGVK 934

Qy     321 cattgc 326
               : |:
Db     935 ACKGGS 940


RESULT   7
CNS017SY
LOCUS       CNS017SY     1101 bp    DNA             GSS       26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN37L08 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL108460
VERSION     AL108460.1  GI:5628764
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Plasmid Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN37L08"
                     /note="end : SP6"
BASE COUNT      254 a    176 c    160 g    152 t    359 others
ORIGIN


  Query Match             4.4%;  Score 43.2;  DB 219;  Length 1101;
  Best Local Similarity   16.0%;  Pred. No. 0.58;
  Matches   66;  Conservative  177;  Mismatches  168;  Indels    1;  Gaps    1;

Qy     278 ctgccgcggcgcgatcggctgagccgcacgtggggagcaagatcattgctaatctagtga 337
           | |: |||| : |    :: |:::|| ::| ||::     :|:    |  :     | :
Db     638 CGGSAGCGGRRKGCAACKSAGMSSCGGRSGSGGSSCCGGKKAKGGVRGGRVCCAGGGASC 697

Qy     338 atcagctggggattaatgtcacccaaaggagcgtcgtcagcactggagcgccggccacga 397
           | :|   ||  |: | : :  ::   : |: : :    |:::    : : :|| :||||
Db     698 ACMAADCGGCCAKMACCSSSSASSSGSSCASTSSSASRGGMVSSCACSGSGSGGASACGA 757

Qy     398 cgaggtctacggcggtgacgtccacgactaccgccccgcagcgaacaagtccatacgggc 457
           :|||| : : ||:|    :| |:|   :  |:: ::||: : :  :|   :|  ::: :
Db     758 SGAGGGSGRGGGSGCCASSGVCGAACSSASCSSASMCGVSSGSSCSASCGSCCGVSSCSA 817

Qy     458 acaatggccgacctgtga-cggctggattggtggcagctaatagtggtgccagcgcggcc 516
           ::|   :::   : : :| |: : :|:  |  ::: :: :|  :    ::: ::::::::
Db     818 VSASSASSVMSKVASAVASCSAVASGMSAGAVSSSCRSSVASSVSAAASVSSSVSSSSSS 877

Qy     517 tcgtctcccacagccgcggcgaaaccaacaggagaagaaaaggcctccgcggcatgtgaa 576
           ::    ::|: :: :: :: :||  |: ::::::|::  :::| :::: ::: : ::|
Db     878 SVVSAAAVSASSSSASSASMAVAAAAAVAVSSVSVASVVSSSSSCSSSSSASVVVSASVA 937

Qy     577 acgagctccgtggcgataaatgcgacacgcccggcgcttcacaacgcctctctcccgcag 636
           |:  :: ::: :::    | :::::|:::: ::::  : : ::::: : :  :   |
Db     938 ASASSVSSSSSSSVSTSSSASVSSVSAVSMSAVVSSSASSSSSVSVSSVVAVAASAAAAA 997

Qy     637 gcgccaacggatggcgttttggcggcagcagtataccagtcggagggcgagg 688
               ||:::  : :     ::::::|:: :   |:: : ::: :: :::::
Db     998 AAAAAAASSSSASAVAVVSASSSSSSASSSSSSSASSSVSSVSSSSCSVVSV 1049


RESULT   8
BE587138
LOCUS       BE587138      577 bp    mRNA            EST       17-AUG-2000
DEFINITION  WHE0517_A09_A17ZR Secale cereale aluminum-stressed root tip cDNA
            library Secale cereale cDNA clone WHE0517_A09_A17, mRNA sequence.
ACCESSION   BE587138
VERSION     BE587138.1  GI:9840170
KEYWORDS    EST.
SOURCE      rye.
  ORGANISM  Secale cereale
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Secale.
REFERENCE   1  (bases 1 to 577)
  AUTHORS   Anderson,O.D., Butler,E., Chao,S., Gustafson,J.P., Han,P.S., Hsia
            ,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J., Seaton,C.L. and
            Tong,J.C.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - Aluminum-stressed root tip cDNA library from rye (Secale
            cereale)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: M13 reversed primer.
FEATURES             Location/Qualifiers
     source          1. .577
                     /organism="Secale cereale"
```

```
                     /cultivar="Blanco"
                     /db_xref="taxon:4550"
                     /clone="WHE0517_A09_A17"
                     /clone_lib="Secale cereale aluminum-stressed root tip cDNA
                     library"
                     /tissue_type="Root tip"
                     /dev_stage="Seedling"
                     /lab_host="E. coli DH12S"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; Seeds
                     were germinated aseptically on filter paper and
                     transferred to a  hydroponic growth system in a growth
                     chamber when the primary root  was 1 cm in length. After a
                     2-day establishment period, seedlings were  subjected to a
                     5 ppm aluminum stress prior to tissue harvest. Plants
                     were grown in an environmental chamber. The tissue, total
                     RNA, and  poly(A) RNA were prepared, and a cDNA library
                     was made (Butler and  Gustafson) at University of Missouri
                     , Columbia. Plasmid DNA  preparations and DNA sequencing
                     were performed in the OD Anderson  lab (all other authors
                     )."
BASE COUNT       97 a    151 c    205 g    124 t
ORIGIN

  Query Match             4.4%;  Score 43;  DB 137;  Length 577;
  Best Local Similarity   48.6%;  Pred. No. 0.61;
  Matches  118;  Conservative    0;  Mismatches  125;  Indels    0;  Gaps    0;

Qy    470 ctgtgacggctggattggtggcagctaatagtggtgccagcgcggcctcgtctcccacag 529
          | |  |||| |   ||| ||||||||     |  ||    |||||  || || |  | |
Db     19 CAGCTACGGAGGTGGTGGCGGCAGCTACGGCGGCGGCGGCGGCGGCTACGGCTACGGCGG 78

Qy    530 ccgcggcgaaaccaacaggagaagaaaaggcctccgcggcatgtgaaacgagctccgtgg 589
          | |||||| |  | || |  || ||  |||| | || |   |||  || ||    | |
Db     79 CGGCGGCGGACGCTACGGAGGAGGAGGAGGCGGCGGCCGTGAGTGCTACAAGTGTGGCGA 138

Qy    590 cgataaatgcgacacgcccggcgcttcacaacgcctctctcccgcaggcgccaacggatg 649
           ||      |  | |    | |    | ||  ||  |     ||  || | | | |||| |
Db    139 GGAGGGCCACATCTCCAGGGACTGCACCCAGGGCGGCGGCGGCGGCGGTGGCTATGGAGG 198

Qy    650 gcgttttggcggcagcagtataccagtcggagggcgaggttcatcagtcgctggagcggc 709
          |   |||| |  || |       |  || ||  | || |  |    |   || | | |
Db    199 CGGCGGTGGCCGTGGCGGCGGCGGCGGTGGCGGCGGCGGCTGCTTCTCCTGCGGCGAGTC 258

Qy    710 tgg 712
          |||
Db    259 TGG 261


RESULT   9
CNS0091P
LOCUS       CNS0091P      925 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL053013
VERSION     AL053013.1  GI:4934461
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 925)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .925
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR19D16"
                     /note="end : TET3"
BASE COUNT      120 a     61 c     61 g    172 t    511 others
ORIGIN

  Query Match             4.4%;  Score 42.8;  DB 219;  Length 925;
  Best Local Similarity   12.8%;  Pred. No. 0.73;
  Matches   45;  Conservative  154;  Mismatches  153;  Indels    0;  Gaps    0;

Qy    369 cgtcgtcagcactggagcgccggccacgacgaggtctacggcggtgacgtccacgactac 428
          :: :|:  || : :| :|:||::|: :: :    :|  | :|:::  :: :: :: :  :
Db    555 SSGSGYGKGCSSGSGBSCSCCSSCSCSSSSCSCCBCCCCCSCSSYCCSSSBSSSKCSSTS 614

Qy    429 cgccccgcagcgaacaagtccatacgggcacaatggccgacctgtgacggctggattggt 488
          ::|:|| : :::     : ::   ::::: :    :::   : : | :::: :: ||::
Db    615 BSCSCCCSSKSVCGTSCSSSSSCSSSSSSTSSSSTSSSTSKSSSSGSSSSSSSSYTTSKS 674

Qy    489 ggcagctaatagtggtgccagcgcggcctcgtctcccacagccgcggcgaaaccaacagg 548
          :  |:  : || |:':    :::::::  :|: ::: : : :: :::|   ::  : ::
Db    675 TSASGSGSWSAGGGSGSTGSTSSSSSSSSTSTSSSSVSSGSKSSTBSSGSBSSSGSSSSS 734

Qy    549 agaagaaaaggcctccgcggcatgtgaaacgagctccgtggcgataaatgcgacacgccc 608
          :    : :: :::: :: |       |   :::::|:::: | :   :| : :::
Db    735 STSSBBSCTSTSSSSSSSYSSSTCSCCTCCCSYSYSSSTSSSSSTSWGSTSGSSSSSVGT 794

Qy    609 ggcgcttcacaacgcctctctcccgcaggcgccaacggatggcgttttggcggcagcagt 668
          ::::  |: |  | |: ||| : :::    : :   ::  :| |  |: |||||  ::
Db    795 SSSSDSTSTCCSCCCYMCTCCSTYBMBCYTSTSCGGSSSSSGKGGVTKCGCGGCGSSSTN 854

Qy    669 ataccagtcggagggcgaggttcatcagtcgctggagcggctggagtccgtc 720
          : |:: :::|: ::  :    : ::: :: : ::  :: : : :::
Db    855 GMBGTSSACSSSSSSCSSSSVSSSSKSSASSSSVSSSGSSGVSSNSSSASKS 906


RESULT  10
CNS010B7
LOCUS       CNS010B7      787 bp    DNA             GSS       26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN03M07 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL098749
VERSION     AL098749.1  GI:5610360
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Plasmid Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 787)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
```

```
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .787
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN03M07"
                     /note="end : T7"
BASE COUNT       52 a    114 c     73 g     19 t    529 others
ORIGIN


  Query Match             4.4%;  Score 42.6;  DB 219;  Length 787;
  Best Local Similarity   2.3%;  Pred. No. 0.81;
  Matches    9;  Conservative  133;  Mismatches  256;  Indels    0;  Gaps    0;

Qy     565 gcggcatgtgaaacgagctccgtggcgataaatgcgacacgcccggcgcttcacaacgcc 624
           : : |   :     :   : :     ::   ::::       : ::  :     : : :
Db     380 KNKCCNNCKNNKNNKNMKKNMNNMNNKMNKNMKKMNNKKNNNMNKKNTMMMNNKMCMMNM 439

Qy     625 tctctcccgcaggcgccaacggatggcgttttggcggcagcagtataccagtcggagggc 684
           ::  :: :  :: :      :  ::::       : :: :   ::   ::::: :
Db     440 KMMNKMNMMNMKNMMNNNNKKNNKKKMNNNNNMNNKNMMNMNNNMKNNNMKKMKNMMNTN 499

Qy     685 gaggttcatcagtcgctggagcggctggagtccgtcataaccaacacgtctcgggttctg 744
           :: |    :   :   ::      ::     ::::: ::    :     :  ::
Db     500 NMKNTNNNNMNNNMMKNKKNMNMMNNNKMNNNKMKMMKMKMMNKNNMMNTMNNMKNNMKN 559

Qy     745 aagttgctccctgacaccattcgaagagaccatgaacaacttctgaatctgggtttagag 804
           : : :      : |:  :::::    : ::: : : : ::    : ::  :|    :
Db     560 NKKNKMMMKNNNTMNAMKNKKMKMNNNNMNMMKNMNMKMNKKNNNNMNMKNMKTMNNNMN 619

Qy     805 gcacagatgacagagctgcagcagagccgtccaacaccgcaaacacagccgagagacaca 864
           :  :         :         : ::   :   :::   ::: :    : ::|  ::
Db     620 NNMNNKNNNNNNNNNMMCMKKNNNCTKNMKNNKMNNNMMKNNNMMMTMNNNCMNMKANTMM 679

Qy     865 agctccgcgaaatcatccgtgtttgagacgtacacccttgttctcattgcggattccctc 924
             ::: |   :   | ::      ::  :        :| : : : : :        :
Db     680 NMNKMMMCMNNMMNNTNMKNNNNMMNNMMNNMNKNNNMMNKTNKGMNKTMNKCNMNNNMM 739

Qy     925 tctcgcaacatcacgaaggggtgaagcgtggtgtgaa 962
             ::    : ::   ::          :: :  |:
Db     740 MNMNNMMNNNNMNMKKNNKKNNMMNKNNNNKKNKMMAM 777


RESULT  11
BE675949
LOCUS       BE675949      504 bp    mRNA            EST       08-SEP-2000
DEFINITION  7f19a02.x1 NCI_CGAP_CLL1 Homo sapiens cDNA clone IMAGE:3295082 3',
            mRNA sequence.
ACCESSION   BE675949
VERSION     BE675949.1  GI:10036490
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 504)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever,
            M.D., Louis M. Staudt, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center.
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            Seq primer: -40UP from Gibco
            High quality sequence stop: 494.
FEATURES             Location/Qualifiers
     source          1. .504
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3295082"
                     /clone_lib="NCI_CGAP_CLL1"
                     /tissue_type="B-cell, chronic lymphotic leukemia"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATTGCTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library is normalized, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT      129 a    154 c    141 g     80 t
ORIGIN


  Query Match             4.3%;  Score 42.4;  DB 138;  Length 504;
  Best Local Similarity   48.7%;  Pred. No. 0.86;
  Matches  115;  Conservative    0;  Mismatches  121;  Indels    0;  Gaps    0;

Qy     538 aaaccaacaggagaagaaaaggcctccgcggcatgtgaaacgagctccgtggcgataaat 597
           ||| ||| |  | || .|||    ||  ||  |    || | |||   ||      || |
Db     269 AAAACAAAACAAAAAACAAATCGGTCAACGTTACCAAAACCAAGCAGAGTCTAACAAACT 328

Qy     598 gcgacacgcccggcgcttcacaacgcctctctcccgcaggcgccaacggatggcgttttg 657
           |  |||  || |  |   ||| | | ||   | || | || ||   | |||||  || ||
Db     329 GTCACAGCCCAGAGGAGGCACGAGGACTGAATGCCACGGGGGCTCCCAGATGGGATTCTG 388

Qy     658 gcggcagcagtataccagtcggagggcgaggttcatcagtcgctggagcggctggagtcc 717
           |  | |  |    |   |   ||   |   |   ||| |   |  | || |  | ||
Db     389 GGAGAATAAAAGGATGTGAGTGAAAACTGCGGAAATCTGAACCAAGTGCAGGCGTGGTTA 448

Qy     718 gtcataaccaacacgtctcgggttctgaagttgctccctgacaccattcgaagaga 773
           |||||||  |    | | || || | || |    |    |  |||| | ||||||
Db     449 ATCATAACACATCATTATTGGTTTGTTAACTGTGACAAATATGCCATACTAAGAGA 504


RESULT  12
BE401994
LOCUS       BE401994      577 bp    mRNA            EST       21-JUL-2000
DEFINITION  CSB003C12F990908 ITEC CSB Wheat Endosperm Library Triticum aestivum
            cDNA clone CSB003C12, mRNA sequence.
ACCESSION   BE401994
VERSION     BE401994.1  GI:9361462
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 577)
  AUTHORS   Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier
            ,S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P.,
            Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P.,
            Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,Y.,
            Pecchioni,N., Qualset,C., Schuch,W., Selvaraj,G., Shariflou,M.,
            Sorrells,M., Warburton,M. and Wenzel,G.
  TITLE     International Triticeae EST Cooperative (ITEC): Production of
            Expressed Sequence Tags for Species of the Triticeae
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Appels R
```

```
           Div. of Plant Industry, CSIRO
           Canberra ACT 2601 AUSTRALIA
           Tel: 61 62 465496
           Fax: 61 62 465000
           Email: rudi@pi.csiro.au
           International Triticeae EST Cooperative (ITEC)
           http://wheat.pw.usda.gov/genome.
FEATURES             Location/Qualifiers
     source          1. .577
                     /organism="Triticum aestivum"
                     /cultivar="Wyuna"
                     /db_xref="taxon:4565"
                     /clone="CSB003C12"
                     /clone_lib="ITEC CSB Wheat Endosperm Library"
                     /tissue_type="endosperm"
                     /dev_stage="8-12 days post anthesis"
                     /lab_host="Escherichia coli SOLR"
                     /note="Vector: Lambda Zap/Bluescript; Site_1: XhoI;
                     Site_2: EcoRI; Plants grown in Phytotron with 18C/13C
                     (day/night) 16 hour light. M13 Reverse sequencing primer
                     used.  1.0 Kbp average insert size."
BASE COUNT       93 a    164 c    224 g     85 t     11 others
ORIGIN


  Query Match             4.3%;  Score 42.4;  DB 167;  Length 577;
  Best Local Similarity   46.0%;  Pred. No. 0.88;
  Matches  127;  Conservative    0;  Mismatches  149;  Indels    0;  Gaps    0;

Qy     386 cgccggccacgacgaggtctacggcggtgacgtccacgactaccgccccgcagcgaacaa 445
           || ||||  |  |     | | |||||||  | || ||     |||  ||  |    | |
Db     289 CGGCGGCTCCCGCCCTAGCGAAGGCGGTGGTGACCGCGGNGGTCGCGGCGGCGGTGGCTA 348

Qy     446 gtccatacgggcacaatggccgacctgtgacggctggattggtggcagctaatagtggtg 505
               | |  |   |  |       | | | |||||  |  |||||| ||      |   |
Db     349 CGGTGGAGGCGGCTATGGCGGTGGCGGCGGCGGCTACAGAGGTGGCGGCGGTGGCTACNG 408

Qy     506 ccagcgcggcctcgtctcccacagccgcggcgaaaccaacaggagaagaaaaggcctccg 565
           |    | |||  |  || || | || | ||||    ||   |  |  |    |||  |||
Db     409 CGGCGGTGGCGGCAGCTACCGCGGCGGTGGCGGCTACAGANGTGGTGGTGGCGGCGGCCG 468

Qy     566 cggcatgtgaaacgagctccgtggcgataaatgcgacacgcccggcgcttcacaacgcct 625
           |  ||   ||| |       | |||  | | ||| ||| |  |||||        | | |
Db     469 CTACAGAGGAAGCAGCGGTGGCGGCCGTGAGTGCTACAAGTGCGGCGAAGAGGGCCACAT 528

Qy     626 ctctcccgcaggcgccaacggatggcgttttggcgg 661
           ||     |   || || | ||  |  ||  | ||
Db     529 CTNCAAGGACTGCCCCCAGGGCNGCNGTNGCNGTGG 564


RESULT  13
BF152676/c
LOCUS       BF152676      390 bp    mRNA            EST       26-OCT-2000
DEFINITION  uz94e04.y1 NCI_CGAP_Te1 Mus musculus cDNA clone IMAGE:3809358 5',
            mRNA sequence.
ACCESSION   BF152676
VERSION     BF152676.1  GI:11034071
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 390)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov

            MGI:1451470
            Seq primer: -40RP from Gibco
            High quality sequence stop: 371.
FEATURES             Location/Qualifiers
     source          1. .390
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3809358"
                     /clone_lib="NCI_CGAP_Te1"
                     /sex="male"
                     /dev_stage="5 months"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.3 kb.  Library constructed by Life
                     Technologies."
BASE COUNT       49 a    150 c    135 g     56 t
ORIGIN


  Query Match             4.3%;  Score 42.2;  DB 145;  Length 390;
  Best Local Similarity   54.1%;  Pred. No. 0.94;
  Matches   86;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy     137 cggtagtggcgacggcgaaaccggggtgtgtgtcctcgggcactgacgcggcgagtagca 196
           ||| || | | || ||    || |||  | ||  || ||||   | |   ||||     |
Db     172 CGGCAGCGCCCACCGCACCGCCTGGGCCTCTGGACTGGGGCTGCGGCTGCGCGACCCCAA 113

Qy     197 gtcataccaatacgacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgc 256
           | | | |  |  |  | || ||||||||||||| |  || |     |  |||| ||  |
Db     112 GGCGTCCAGAAGCAGCCTCCGCCGCTGCTGCAGCAGCTCCTGACGCTGCTGTCTCCTCCT 53

Qy     257 cagccaaggcagccgcgcctcctgccgcggcgcgatcgg 295
           ||    || || ||  |  || |||| || |  |||
Db      52 GTGCAGCAGCTGCTGCCACAGCTCCCGCCCCGTGCCCGG 14


RESULT  14
BE595680
LOCUS       BE595680      391 bp    mRNA            EST       18-AUG-2000
DEFINITION  PI1_55_C06.g1_A002 Pathogen induced 1 (PI1) Sorghum bicolor cDNA,
            mRNA sequence.
ACCESSION   BE595680
VERSION     BE595680.1  GI:9850753
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 391)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Dean,R., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: pathogen-induced plants
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: PolyTMix
            High quality sequence start: 2
            High quality sequence stop: 260
            POLYA=No.
```

```
FEATURES             Location/Qualifiers
     source          1. .391
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Pathogen induced 1 (PI1)"
                     /note="Organ: Anthracnose-infected leaves from
                     two-week-old sorghum plants 48 hr after inoculation;
                     Vector: pBluescript II from Lambda Zap II; Site_1: XhoI;
                     Site_2: EcoRI; Two-week-old sorghum plants (BTX 623
                     cultivar) were infected with pathogen (isolate FRM42I of
                     Colletotrichum graminicola, which is a sorghum isolate).
                     RNA was prepared from infected leaves harvested from 45
                     seedlings  48 hours after inoculation.  Note: young
                     seedlings (2 weeks old) exhibit juvenile resistant
                     reaction, which is an incompatible interaction. As they
                     grow older (4 weeks or older), plants resume susceptiblity
                     to anthracnose disease. The library was made from poly-A
                     RNA in the cloning vector lambda ZAP II. Clones to be
                     sequenced were prepared by mass excision. WARNING: While
                     most or all ESTs are expected to derive from the host
                     plant, no effort was made to eliminate ESTs deriving from
                     the pathogen."
BASE COUNT       53 a     99 c    161 g     78 t
ORIGIN


  Query Match             4.3%;  Score 41.8;  DB 137;  Length 391;
  Best Local Similarity   49.3%;  Pred. No. 1.2;
  Matches  109;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps    0;

Qy    210 gacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaaggcagc 269
          | || | |  || |  ||  |  || |  |  || | ||    | | | || | ||  |
Db     12 GGCTACGGAGGCGGTGGCGGCGGCTACGGCGGCCGCCGTGAGGGTGGCGGCTACGGAGGA 71

Qy    270 cgcgcctcctgccgcggcgcgatcggctgagccgcacgtggggagcaagatcattgctaa 329
           ||| |  | ||  ||||| | || ||   |  | | | ||  |   ||    | ||||
Db     72 GGCGGCGGCGGCTACGGCGGGCTCCGCGAGGGAGGAGGCGGCTACGGAGGCGGTGGCTAC 131

Qy    330 tctagtgaatcagctggggattaatgtcacccaaaggagcgtcgtcagcactggagcgcc 389
            | | |     |  ||  |     ||| ||   |||   |  |     |  |  ||| |
Db    132 GGTGGCGGCGGCGGCGGCTACGGCGGTCGCCGTGAGGGCGGCGGCGGCTATGGCGGCGGC 191

Qy    390 ggccacgacgaggtctacggcggtgacgtccacgactaccg 430
          |||  | |||  | | || ||||||||  |  || | || |
Db    192 GGCGGCTACGGTGGCAACCGCGGTGACTCCGGCGGCAACTG 232


RESULT  15
AW747489
LOCUS       AW747489      488 bp    mRNA            EST       19-JUL-2000
DEFINITION  WS1_68_D09.g1_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA,
            mRNA sequence.
ACCESSION   AW747489
VERSION     AW747489.1  GI:7661227
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 488)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: water-stressed plants
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: T7
            High quality sequence start: 4
            High quality sequence stop: 488
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .488
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Water-stressed 1 (WS1)"
                     /note="Organ: Mix of 5-week old plants on days 7 & 8 after
                     water was withheld; Vector: Lambda Zap; Site_1: XhoI;
                     Site_2: EcoRI; The library was made from poly-A RNA in the
                     cloning vector lambda ZAP II.  Clones to be sequenced were
                     prepared by mass excision."
BASE COUNT       73 a    138 c    189 g     88 t
ORIGIN


  Query Match             4.3%;  Score 41.8;  DB 120;  Length 488;
  Best Local Similarity   49.3%;  Pred. No. 1.2;
  Matches  109;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps    0;

Qy    210 gacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaaggcagc 269
          | || | |  || |  ||  |  || |  |  |||| ||    | | | || | ||  |
Db    113 GGCTACGGAGGCGGTGGCGGCGGCTACGGCGGCCCCCGTGAGGGTGGCGGCTACGGAGGA 172

Qy    270 cgcgcctcctgccgcggcgcgatcggctgagccgcacgtggggagcaagatcattgctaa 329
           ||| |  | ||  ||||| |  | ||   |  | | | ||  |   ||    | ||||
Db    173 GGCGGCGGCGGCTACGGCGGGCCCCGCGAGGGAGGAGGCGGCTACGGAGGCGGTGGCTAC 232

Qy    330 tctagtgaatcagctggggattaatgtcacccaaaggagcgtcgtcagcactggagcgcc 389
            | | |     |  ||  |     ||| ||   |||   |  |     |  |  ||| |
Db    233 GGTGGCGGCGGCGGCGGCTACGGCGGTCCCCGTGAGGGCGGCGGCGGCTATGGCGGCGGC 292

Qy    390 ggccacgacgaggtctacggcggtgacgtccacgactaccg 430
          |||  | |||  | | || ||||||||  |  || | || |
Db    293 GGCGGCTACGGTGGCAACCGCGGTGACTCCGGCGGCAACTG 333


Search completed: September 20, 2001, 23:27:52
Job time: 25532 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: September 21, 2001, 02:16:14 ; Search time 164.23 Seconds
(without alignments)
1125.054 Million cell updates/sec

Title: US-09-138-735-1_COPY_1232_2207
Perfect score: 976
Sequence: 1 caggtacagcgtaacggctt.....tgtgaacgaggccattatgt 976

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn1_7/ptodata/1/ina/5A_COMB.seq:*
2: /cgn1_7/ptodata/1/ina/5B_COMB.seq:*
3: /cgn1_7/ptodata/1/ina/6A_COMB.seq:*
4: /cgn1_7/ptodata/1/ina/6B_COMB.seq:*
5: /cgn1_7/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn1_7/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	976	100.0	3402	1	US-08-480-917-1	Sequence 1, Appli
	2	39.2	4.0	2790	2	US-08-718-661-1	Sequence 1, Appli
c	3	38.2	3.9	2943	1	US-08-042-747A-7	Sequence 7, Appli
	4	37.6	3.9	2115	2	US-08-474-379C-60	Sequence 60, Appl
	5	37.6	3.9	2115	3	US-09-146-249A-60	Sequence 60, Appl
	6	37.6	3.9	2115	3	US-08-206-188B-60	Sequence 60, Appl
	7	36.8	3.8	3777	3	US-09-121-321-15	Sequence 15, Appl
	8	36.8	3.8	3777	4	US-08-933-803A-15	Sequence 15, Appl
c	9	36.6	3.8	203	4	US-09-043-303-7	Sequence 7, Appli
	10	36.6	3.8	913	1	US-08-217-327-3	Sequence 3, Appli
	11	36.6	3.8	913	1	US-07-885-970A-3	Sequence 3, Appli
	12	36.6	3.8	913	1	US-08-298-687A-3	Sequence 3, Appli
	13	36.6	3.8	913	1	US-08-530-797-2	Sequence 2, Appli
	14	36.6	3.8	913	1	US-08-298-829-3	Sequence 3, Appli
	15	36.6	3.8	913	2	US-08-787-335-2	Sequence 2, Appli
	16	36.6	3.8	1984	1	US-07-885-970A-25	Sequence 25, Appl
	17	36.6	3.8	1985	1	US-08-298-687A-25	Sequence 25, Appl
	18	36.6	3.8	1985	1	US-08-298-829-25	Sequence 25, Appl
c	19	36.2	3.7	336	2	US-07-814-220-4	Sequence 4, Appli
c	20	36.2	3.7	336	2	US-07-812-421-4	Sequence 4, Appli
c	21	36	3.7	2580	3	US-09-050-863-2	Sequence 2, Appli
	22	36	3.7	5452	2	US-09-130-114-1	Sequence 1, Appli
c	23	36	3.7	9600	4	US-08-910-647-1	Sequence 1, Appli
c	24	36	3.7	10596	1	US-07-884-811-15	Sequence 15, Appl
c	25	36	3.7	10596	1	US-07-885-971-15	Sequence 15, Appl
c	26	36	3.7	10596	1	US-08-087-783A-15	Sequence 15, Appl
c	27	36	3.7	10596	1	US-08-194-088B-15	Sequence 15, Appl
c	28	36	3.7	10596	2	US-08-194-087-15	Sequence 15, Appl
c	29	36	3.7	10596	5	PCT-US93-04648-15	Sequence 15, Appl
c	30	35.2	3.6	3376	1	US-08-320-559-29	Sequence 29, Appl
c	31	35.2	3.6	3376	3	US-08-545-860D-29	Sequence 29, Appl
c	32	35.2	3.6	3376	5	PCT-US94-04496-29	Sequence 29, Appl
c	33	35	3.6	168	1	US-08-469-802B-4	Sequence 4, Appli
c	34	35	3.6	168	2	US-08-267-803B-4	Sequence 4, Appli
c	35	35	3.6	171	1	US-08-469-802B-5	Sequence 5, Appli
c	36	35	3.6	171	2	US-08-267-803B-5	Sequence 5, Appli
c	37	35	3.6	195	1	US-08-469-802B-2	Sequence 2, Appli
c	38	35	3.6	195	2	US-08-267-803B-2	Sequence 2, Appli
c	39	35	3.6	234	1	US-08-469-802B-3	Sequence 3, Appli
c	40	35	3.6	234	2	US-08-267-803B-3	Sequence 3, Appli
c	41	35	3.6	336	2	US-07-814-220-3	Sequence 3, Appli
c	42	35	3.6	336	2	US-07-812-421-3	Sequence 3, Appli
	43	34.8	3.6	1299	1	US-07-661-610C-11	Sequence 11, Appl
c	44	34.8	3.6	2001	1	US-08-674-168-24	Sequence 24, Appl
c	45	34.8	3.6	2001	3	US-08-985-908-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-08-480-917-1
; Sequence 1, Application US/08480917
; Patent No. 5820864
; GENERAL INFORMATION:
; APPLICANT: PARANHOS-BACCALA, Glaucia
; APPLICANT: LESENECHAL, Mylene
; APPLICANT: JOLIVET, Michel
; TITLE OF INVENTION: NEW TRYPANOSOMA CRUZI ANTIGEN, AND GENE
; TITLE OF INVENTION: ENCODING THE LATTER; THEIR APPLICATION TO THE DETECTION OF
; TITLE OF INVENTION: CHAGAS DISEASE
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oliff & Berridge
; STREET: 700 South Washington Street, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,917
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 36400
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3402 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-480-917-1

Query Match 100.0%; Score 976; DB 1; Length 3402;
Best Local Similarity 100.0%; Pred. No. 6.9e-247;
Matches 976; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caggtacagcgtaacggcttttgcttcaatcgtacagccgacggtagctgcgtcctggct 60

```
Db   1232 CAGGTACAGCGTAACGGCTTTTGCTTCAATCGTACAGCCGACGGTAGCTGCGTCCTGGCT 1291

Qy     61 gacatgtcgattcgattgacgatcttccatctccggtcctcccgcagggaagaacagcag 120
Db   1292 GACATGTCGATTCGATTGACGATCTTCCATCTCCGGTCCTCCCGCAGGGAAGAACAGCAG 1351

Qy    121 ccaggccaaaaaacatcggtagtggcgacggcgaaaccggggtgtgtgtcctcgggcact 180
Db   1352 CCAGGCCAAAAAACATCGGTAGTGGCGACGGCGAAACCGGGGTGTGTGTCCTCGGGCACT 1411

Qy    181 gacgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatcccctgcatca 240
Db   1412 GACGCGGCGAGTAGCAGTCATACCAATACGACTTCTGCCGCTGCTGCATCCCCTGCATCA 1471

Qy    241 ccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggctgag 300
Db   1472 CCCCCTGTTTCAGCGCCAGCCAAGGCAGCCGCGCCTCCTGCCGCGGCGCGATCGGCTGAG 1531

Qy    301 ccgcacgtggggagcaagatcattgctaatctagtgaatcagctggggattaatgtcacc 360
Db   1532 CCGCACGTGGGGAGCAAGATCATTGCTAATCTAGTGAATCAGCTGGGGATTAATGTCACC 1591

Qy    361 caaaggagcgtcgtcagcactggagcgccggccacgacgaggtctacggcggtgacgtcc 420
Db   1592 CAAAGGAGCGTCGTCAGCACTGGAGCGCCGGCCACGACGAGGTCTACGGCGGTGACGTCC 1651

Qy    421 acgactaccgccccgcagcgaacaagtccatacgggcacaatggccgacctgtgacggct 480
Db   1652 ACGACTACCGCCCCGCAGCGAACAAGTCCATACGGGCACAATGGCCGACCTGTGACGGCT 1711

Qy    481 ggattggtggcagctaatagtggtgccagcgcggcctcgtctcccacagccgcggcgaaa 540
Db   1712 GGATTGGTGGCAGCTAATAGTGGTGCCAGCGCGGCCTCGTCTCCCACAGCCGCGGCGAAA 1771

Qy    541 ccaacaggagaagaaaaggcctccgcggcatgtgaaacgagctccgtggcgataaatgcg 600
Db   1772 CCAACAGGAGAAGAAAAGGCCTCCGCGGCATGTGAAACGAGCTCCGTGGCGATAAATGCG 1831

Qy    601 acacgcccggcgcttcacaacgcctctctcccgcaggcgccaacggatggcgttttggcg 660
Db   1832 ACACGCCCGGCGCTTCACAACGCCTCTCTCCCGCAGGCGCCAACGGATGGCGTTTTGGCG 1891

Qy    661 gcagcagtataccagtcggagggcgaggttcatcagtcgctggagcggctggagtccgtc 720
Db   1892 GCAGCAGTATACCAGTCGGAGGGCGAGGTTCATCAGTCGCTGGAGCGGCTGGAGTCCGTC 1951

Qy    721 ataaccaacacgtctcgggttctgaagttgctccctgacaccattcgaagagaccatgaa 780
Db   1952 ATAACCAACACGTCTCGGGTTCTGAAGTTGCTCCCTGACACCATTCGAAGAGACCATGAA 2011

Qy    781 caacttctgaatctgggtttagaggcacagatgacagagctgcagcagagccgtccaaca 840
Db   2012 CAACTTCTGAATCTGGGTTTAGAGGCACAGATGACAGAGCTGCAGCAGAGCCGTCCAACA 2071

Qy    841 ccgcaaacacagccgagagacacaagctccgcgaaatcatccgtgtttgagacgtacacc 900
Db   2072 CCGCAAACACAGCCGAGAGACACAAGCTCCGCGAAATCATCCGTGTTTGAGACGTACACC 2131

Qy    901 cttgttctcattgcggattccctctctcgcaacatcacgaaggggtgaagcgtggtgtg 960
Db   2132 CTTGTTCTCATTGCGGATTCCCTCTCTCGCAACATCACGAAGGGGGTGAAGCGTGGTGTG 2191

Qy    961 aacgaggccattatgt 976
Db   2192 AACGAGGCCATTATGT 2207


RESULT   2
US-08-718-661-1
; Sequence 1, Application US/08718661
; Patent No. 5876972
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Nucleic acid molecules coding for mammalian
;    TITLE OF INVENTION:  tumor suppressor proteins and methods for their isolation
;    NUMBER OF SEQUENCES:  15
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/718,661
;      FILING DATE:
;      CLASSIFICATION:  530
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2790 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  542..2545
US-08-718-661-1


  Query Match             4.0%;  Score 39.2;  DB 2;  Length 2790;
  Best Local Similarity   54.1%;  Pred. No. 0.24;
  Matches   80;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy    137 cggtagtggcgacggcgaaaccggggtgtgtgtcctcgggcactgacgcggcgagtagca 196
Db   2114 CAGCAGCAGCCACCGCCACTGCCGCCGCCACCACCGCTGCCACTGCCACAGCCACTGCCA 2173

Qy    197 gtcataccaatacgacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgc 256
Db   2174 GTGCCACAGCCACTACCACAGCCACAGATGCAGCCACAGTTTCAGTTGCAGATCCAGCCC 2233

Qy    257 cagccaaggcagccgcgcctcctgccgc 284
Db   2234 CAGATGCAGCTACCACAGCTGCTGCCGC 2261


RESULT   3
US-08-042-747A-7/c
; Sequence 7, Application US/08042747A
; Patent No. 5487969
;  GENERAL INFORMATION:
;    APPLICANT:  Eberle, Richard
;    APPLICANT:  Black, Darla
;    APPLICANT:  Scinicariello, Franco
;    APPLICANT:  Hilliard, Julia K.
;    TITLE OF INVENTION:  Cloning and Amplification of Monkey B
;    TITLE OF INVENTION:  Virus Genes
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cox & Smith Incorporated
;      STREET:  112 East Pecan Street, Suite 2000
;      CITY:  San Antonio
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  78205
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/042,747A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Haymond, W. Bradley
```

```
;       REGISTRATION NUMBER:  35186
;       REFERENCE/DOCKET NUMBER:  S-0072.179
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  210-554-5500
;       TELEFAX:  210-226-8395
;       TELEX:  767609
;   INFORMATION FOR SEQ ID NO:  7:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2943 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  Genomic DNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  87..2744
US-08-042-747A-7


  Query Match             3.9%;  Score 38.2;  DB 1;  Length 2943;
  Best Local Similarity   47.7%;  Pred. No. 0.45;
  Matches  112;  Conservative    0;  Mismatches  123;  Indels    0;  Gaps    0;

Qy    209 cgacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaaggcag 268
          || | ||  |||    |  | ||  |  |    |  | | |      ||| | |    ||| |
Db    255 CGTCCTCGTCCGGGACCTCCTCGTCCTCGGGAAGCGCGGGCGTGGCGGCGGGGTCGGCGG 196

Qy    269 ccgcgcctcctgccgcggcgcgatcggctgagccgcacgtggggagcaagatcattgcta 328
           || | |  | ||||| ||| |  |||||   |||  || || || || |    |  || |
Db    195 TCGGGGCCGCCGCCGCCGCGGGGGCGGCTCGTCCGGCCGCGGCGATCACGGCGAGGGCGA 136

Qy    329 atctagtgaatcagctggggattaatgtcacccaaaggagcgtcgtcagcactggagcgc 388
                |     |    | |   | ||          | | | |  || | | ||
Db    135 GGAGGACGGGGAGGGGAAGAAAGGAGGGGGGGGTGCCGCGAGGCCGCATCCCGGGGCGCG 76

Qy    389 cggccacgacgaggtctacggcggtgacgtccacgactaccgccccgcagcgaac 443
          ||| | | ||| || | ||  ||  | |||| ||  |  ||  || || |||| |
Db     75 CGGGCGCTACGCGGGCGACTACGCGGGCGTCGACTCCGGCCCGCCGGCCGCGAGC 21


RESULT   4
US-08-474-379C-60
; Sequence 60, Application US/08474379C
; Patent No. 5977305
;  GENERAL INFORMATION:
;    APPLICANT:  Wigler, Michael H.
;    APPLICANT:  Colicelli, John J.
;    TITLE OF INVENTION:  CLONING BY COMPLEMENTATION AND RELATED
;    TITLE OF INVENTION:  PROCESSES
;    NUMBER OF SEQUENCES:  88
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  233 South Wacker Drive/6300 Sears Tower
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/474,379C
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/511,715
;      FILING DATE:  20-APR-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/206,188
;      FILING DATE:  01-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/688,352
;      FILING DATE:  19-APR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  27866/32771
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 474-6300
;      TELEFAX:  (312) 474-0448
;  INFORMATION FOR SEQ ID NO:  60:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2115 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  762..2115
US-08-474-379C-60


  Query Match             3.9%;  Score 37.6;  DB 2;  Length 2115;
  Best Local Similarity   63.0%;  Pred. No. 0.59;
  Matches   58;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy    212 cttctgccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaaggcagccg 271
          || |  |||| || ||  |||  ||  || |||| |   ||| | |||||   || ||||
Db    577 CTGCCCCCGCCGCCGCCACCCTCGCCCCAGCCCCAGCCCCAGTGTCAGCCGCCGCCGCCG 636

Qy    272 cgcctcctgccgcggcgcgatcggctgagccg 303
          | || |||||||| || |   | || |  | |
Db    637 CCCCCCCTGCCGCCGCCCCCGCCGCCGCCCGG 668


RESULT   5
US-09-146-249A-60
; Sequence 60, Application US/09146249A
; Patent No. 6069240
;  GENERAL INFORMATION:
;    APPLICANT:  Wigler, Michael H.
;    APPLICANT:  Colicelli, John J.
;    TITLE OF INVENTION:  Cloning by Complementation and Related
;    TITLE OF INVENTION:  Processes
;    NUMBER OF SEQUENCES:  85
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/146,249A
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/511,715
;      FILING DATE:  20-APR-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312-474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  60:
;    SEQUENCE CHARACTERISTICS:
```

```
;      LENGTH:  2115 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  762..2115
US-09-146-249A-60


  Query Match             3.9%;  Score 37.6;  DB 3;  Length 2115;
  Best Local Similarity   63.0%;  Pred. No. 0.59;
  Matches   58;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy    212 cttctgccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaaggcagccg 271
          || |  |||| || ||  |||  ||  || |||| |   ||| | |||||   || ||||
Db    577 CTGCCCCCGCCGCCGCCACCCTCGCCCCAGCCCCAGCCCCAGTGTCAGCCGCCGCCGCCG 636

Qy    272 cgcctcctgccgcggcgcgatcggctgagccg 303
          | || |||||||| || |   | || |  | |
Db    637 CCCCCCCTGCCGCCGCCCCCGCCGCCGCCCGG 668


RESULT   6
US-08-206-188B-60
; Sequence 60, Application US/08206188B
; Patent No. 6100025
;  GENERAL INFORMATION:
;    APPLICANT:  Wigler, Michael H.
;    APPLICANT:  Colicelli, John J.
;    TITLE OF INVENTION:  Cloning by Complementation and Related
;    TITLE OF INVENTION:  Processes
;    NUMBER OF SEQUENCES:  84
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM!
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/206,188B
;      FILING DATE:  01-MAR-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/511,715
;      FILING DATE:  20-APR-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36107
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312-474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  60:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2115 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  762..2115
US-08-206-188B-60


  Query Match             3.9%;  Score 37.6;  DB 3;  Length 2115;
  Best Local Similarity   63.0%;  Pred. No. 0.59;
  Matches   58;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy    212 cttctgccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaaggcagccg 271
          || |  |||| || ||  |||  ||  || |||| |   ||| | |||||   || ||||
Db    577 CTGCCCCCGCCGCCGCCACCCTCGCCCCAGCCCCAGCCCCAGTGTCAGCCGCCGCCGCCG 636

Qy    272 cgcctcctgccgcggcgcgatcggctgagccg 303
          | || |||||||| || |   | || |  | |
Db    637 CCCCCCCTGCCGCCGCCCCCGCCGCCGCCCGG 668


RESULT   7
US-09-121-321-15
; Sequence 15, Application US/09121321
; Patent No. 6090783
;  GENERAL INFORMATION:
;    APPLICANT:  Saiga, Akihiko
;    APPLICANT:  Orita, Satoshi
;    APPLICANT:  Igarashi, Hisanaga
;    APPLICANT:  Okumura, Kouichi
;    APPLICANT:  Sakaguchi, Gaku
;    TITLE OF INVENTION:  DNA MOLECULE RELATING TO SUPPRESSION OF
;    TITLE OF INVENTION:  GENE EXPRESSION AND NOVEL PROTEIN
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FISH & NEAVE
;      STREET:  1251 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10020
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/121,321
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/933,803
;      FILING DATE:  19-SEP-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Haley, James F.
;      REGISTRATION NUMBER:  27,794
;      REFERENCE/DOCKET NUMBER:  SHGN-12CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-596-9000
;      TELEFAX:  212-596-9090
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3777 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  139..2151
US-09-121-321-15


  Query Match             3.8%;  Score 36.8;  DB 3;  Length 3777;
  Best Local Similarity   69.4%;  Pred. No. 1.1;
  Matches   50;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy    231 ccctgcatcacccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcg 290
          ||| ||  || |||| |   ||||||||||||  ||||||||  || | | ||| || |
Db   1355 CCCCGCCACAGCCCCAGCCCCAGCGCCAGCCACCGCAGCCGCAGCTGCAGTCGCAGCCCC 1414
```

```
Qy     291 atcggctgagcc 302
           | |  | |||||
Db    1415 AGCCCCAGAGCC 1426


RESULT   8
US-08-933-803A-15
; Sequence 15, Application US/08933803A
; Patent No. 6218522
;  GENERAL INFORMATION:
;    APPLICANT:  Saiga, Akihiko
;    APPLICANT:  Orita, Satoshi
;    APPLICANT:  Igarashi, Hisanaga
;    APPLICANT:  Okumura, Kouichi
;    APPLICANT:  Sakaguchi, Gaku
;    TITLE OF INVENTION:  DNA MOLECULE RELATING TO SUPPRESSION OF
;    TITLE OF INVENTION:  GENE EXPRESSION AND NOVEL PROTEIN
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FISH & NEAVE
;      STREET:  1251 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10020
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/933,803A
;      FILING DATE:  19-SEP-1997
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Haley, James F.
;      REGISTRATION NUMBER:  27,794
;      REFERENCE/DOCKET NUMBER:  SHGN-12CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-596-9000
;      TELEFAX:  212-596-9090
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3777 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  139..2151
US-08-933-803A-15


  Query Match             3.8%;  Score 36.8;  DB 4;  Length 3777;
  Best Local Similarity   69.4%;  Pred. No. 1.1;
  Matches   50;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy     231 ccctgcatcacccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgcg 290
           ||| ||  || |||| |   ||||||||||||  ||||||||  || | | ||| || | |
Db    1355 CCCCGCCACAGCCCCAGCCCCAGCGCCAGCCACCGCAGCCGCAGCTGCAGTCGCAGCCCC 1414

Qy     291 atcggctgagcc 302
           | |  | |||||
Db    1415 AGCCCCAGAGCC 1426


RESULT   9
US-09-043-303-7/c
; Sequence 7, Application US/09043303
; Patent No. 6251589
; GENERAL INFORMATION:
;  APPLICANT: TSUJI, Shoji
;  APPLICANT: SANPEI, Kazujiro
;  TITLE OF INVENTION: Method for Diagnosing Spinocerebellar Alaxia Type 2 and
;  TITLE OF INVENTION: Primers Therefor
;  FILE REFERENCE: 0760-0241P
;  CURRENT APPLICATION NUMBER: US/09/043,303
;  CURRENT FILING DATE: 1998-05-18
;  EARLIER APPLICATION NUMBER: PCT/JP96/01999
;  EARLIER FILING DATE: 1996-07-18
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
;   LENGTH: 203
;   TYPE: DNA
;   ORGANISM: p-2093 plasmid
US-09-043-303-7


  Query Match             3.8%;  Score 36.6;  DB 4;  Length 203;
  Best Local Similarity   57.4%;  Pred. No. 0.55;
  Matches   66;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy     184 gcggcgagtagcagtcataccaatacgacttctgccgctgctgcatcccctgcatcaccc 243
           ||   ||||  | || || |   | |  || |||| ||||||||  |  ||||  |  |
Db     202 GCCCAGAGTTTCCGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 143

Qy     244 cctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggctg 298
           |||  | | ||  | ||    || || ||  || |||| || ||     | ||||
Db     142 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 88


RESULT  10
US-08-217-327-3
; Sequence 3, Application US/08217327
; Patent No. 5474925
;  GENERAL INFORMATION:
;    APPLICANT:  John, Maliyakal E
;    APPLICANT:  Barton, Kenneth A
;    TITLE OF INVENTION:  Immobilized Proteins in Cotton Fiber
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Quarles and Brady
;      STREET:  P.O. Box 2113
;      CITY:  Madison
;      STATE:  WI
;      COUNTRY:  USA
;      ZIP:  53701-2113
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/217,327
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/812,233
;      FILING DATE:  19-DEC-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seay, Nicholas J
;      REGISTRATION NUMBER:  27,386
;      REFERENCE/DOCKET NUMBER:  1122990831
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  608-251-5000
;      TELEFAX:  608-251-9166
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  913 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
```

```
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Gossypium hirsutum
;      STRAIN:  Coker 312
;      TISSUE TYPE:  Fiber cells
;    IMMEDIATE SOURCE:
;      CLONE:  H6
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  72..716
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  72..147
US-08-217-327-3


  Query Match             3.8%;  Score 36.6;  DB 1;  Length 913;
  Best Local Similarity   58.9%;  Pred. No. 0.85;
  Matches   63;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy     198 tcataccaatacgacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgcc 257
           ||   ||| | |  || || | ||  || || | ||||| || || ||| | || | ||
Db     320 TCCACCCAGTTCTCCTCCTCCTGCAACTCCACCACCTGCTTCTCCTCCTCCTGCAACTCC 379

Qy     258 agccaaggcagccgcgcctcctgccgcggcgcgatcggctgagccgc 304
           | |    ||  |  | |||||||||| |  | | | | |||   || |
Db     380 ACCTCCAGCTTCTCCACCTCCTGCCACTCCTCCACCAGCTTCTCCAC 426


RESULT  11
US-07-885-970A-3
; Sequence 3, Application US/07885970A
; Patent No. 5495070
;  GENERAL INFORMATION:
;    APPLICANT:  John, Maliyakal E.
;    TITLE OF INVENTION:  GENETICALLY ENGINEERING COTTON
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nicholas J. Seay, Quarles & Brady
;      STREET:  P.O. Box 2113, First Wisconsin Plaza
;      CITY:  Madison
;      STATE:  Wisconsin
;      COUNTRY:  USA
;      ZIP:  53701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Microsoft Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/885,970A
;      FILING DATE:  19920518
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/617,239
;      FILING DATE:  21-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/253,243
;      FILING DATE:  04-OCT-1988
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seay, Nicholas J.
;      REGISTRATION NUMBER:  27,386
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (608) 283-2478
;      TELEFAX:  (608) 251-5139
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  913 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Gossypium hirsutum
;      STRAIN:  Coker 312
;      DEVELOPMENTAL STAGE:  15 day old fiber cells
;      TISSUE TYPE:  fiber cells
;    IMMEDIATE SOURCE:
;      LIBRARY:  CKFB15A1
;      CLONE:  H6
US-07-885-970A-3


  Query Match             3.8%;  Score 36.6;  DB 1;  Length 913;
  Best Local Similarity   58.9%;  Pred. No. 0.85;
  Matches   63;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy     198 tcataccaatacgacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgcc 257
           ||   ||| | |  || || | ||  || || | ||||| || || ||| | || | ||
Db     320 TCCACCCAGTTCTCCTCCTCCTGCAACTCCACCACCTGCTTCTCCTCCTCCTGCAACTCC 379

Qy     258 agccaaggcagccgcgcctcctgccgcggcgcgatcggctgagccgc 304
           | |    ||  |  | |||||||||| |  | | | | |||   || |
Db     380 ACCTCCAGCTTCTCCACCTCCTGCCACTCCTCCACCAGCTTCTCCAC 426


RESULT  12
US-08-298-687A-3
; Sequence 3, Application US/08298687A
; Patent No. 5521078
;  GENERAL INFORMATION:
;    APPLICANT:  John, Maliyakal E.
;    TITLE OF INVENTION:  GENETICALLY ENGINEERING COTTON
;    TITLE OF INVENTION:  PLANTS FOR ALTERED FIBER
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nicholas J. Seay, Quarles & Brady
;      STREET:  P.O. Box 2113, First Wisconsin Plaza
;      CITY:  Madison
;      STATE:  Wisconsin
;      COUNTRY:  USA
;      ZIP:  53701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Microsoft Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/298,687A
;      FILING DATE:
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/617,239
;      FILING DATE:  21-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/253,243
;      FILING DATE:  04-OCT-1988
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seay, Nicholas J.
;      REGISTRATION NUMBER:  27,386
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (608) 283-2478
;      TELEFAX:  (608) 251-5139
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  913 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Gossypium hirsutum
```

```
;      STRAIN:  Coker 312
;      DEVELOPMENTAL STAGE:  15 day old fiber cells
;      TISSUE TYPE:  fiber cells
;    IMMEDIATE SOURCE:
;      LIBRARY:  CKFB15A1
;      CLONE:  H6
US-08-298-687A-3


  Query Match             3.8%;  Score 36.6;  DB 1;  Length 913;
  Best Local Similarity   58.9%;  Pred. No. 0.85;
  Matches   63;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy    198 tcataccaatacgacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgcc 257
          ||   ||| | |  || || | ||  || || | ||||| || || |||  | || | ||
Db    320 TCCACCCAGTTCTCCTCCTCCTGCAACTCCACCACCTGCTTCTCCTCCTCCTGCAACTCC 379

Qy    258 agccaaggcagccgcgcctcctgccgcggcgcgatcggctgagccgc 304
          | |    ||  |  | |||||||||| |  | | | | |||   || |
Db    380 ACCTCCAGCTTCTCCACCTCCTGCCACTCCTCCACCAGCTTCTCCAC 426


RESULT  13
US-08-530-797-2
; Sequence 2, Application US/08530797
; Patent No. 5597718
;  GENERAL INFORMATION:
;    APPLICANT:  John, Maliyakal E.
;    APPLICANT:  Umbeck, Paul F.
;    APPLICANT:  Brill, Winston J.
;    TITLE OF INVENTION:  GENETICALY ENGINEERED COTTON PLANTS
;    TITLE OF INVENTION:  FOR ALTERED FIBER
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Quarles and Brady
;      STREET:  P.O BOX 2113
;      STREET:  FIRST WISCONSIN PLAZA
;      CITY:  MADISON
;      STATE:  WISCONSIN
;      COUNTRY:  U.S.A.
;      ZIP:  53701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette - 3.50 inch, 800Kb storage
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Macintosh
;      SOFTWARE:  Microsoft Word 4.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/530,797
;      FILING DATE:  20-SEP-1995
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/617,239
;      FILING DATE:  21-NOV-90
;      APPLICATION NUMBER:  US 07/253,243
;      FILING DATE:  04-OCT-88
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nicholas J. Seay
;      REGISTRATION NUMBER:  27,386
;      REFERENCE/DOCKET NUMBER:  1122990245
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  913 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  no
;    ANTI-SENSE:  no
;    ORIGINAL SOURCE:
;      ORGANISM:  Gossypium hirsutum
;      STRAIN:  Coker 312
;      DEVELOPMENTAL STAGE:  15 day old fiber cells
;      TISSUE TYPE:  fiber cells
;    IMMEDIATE SOURCE:
;      LIBRARY:  CKFB15A1
;      CLONE:  H6
US-08-530-797-2


  Query Match             3.8%;  Score 36.6;  DB 1;  Length 913;
  Best Local Similarity   58.9%;  Pred. No. 0.85;
  Matches   63;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy    198 tcataccaatacgacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgcc 257
          ||   ||| | |  || || | ||  || || | ||||| || || |||  | || | ||
Db    320 TCCACCCAGTTCTCCTCCTCCTGCAACTCCACCACCTGCTTCTCCTCCTCCTGCAACTCC 379

Qy    258 agccaaggcagccgcgcctcctgccgcggcgcgatcggctgagccgc 304
          | |    ||  |  | |||||||||| |  | | | | |||   || |
Db    380 ACCTCCAGCTTCTCCACCTCCTGCCACTCCTCCACCAGCTTCTCCAC 426


RESULT  14
US-08-298-829-3
; Sequence 3, Application US/08298829
; Patent No. 5620882
;  GENERAL INFORMATION:
;    APPLICANT:  John, Maliyakal E.
;    TITLE OF INVENTION:  GENETICALLY ENGINEERING COTTON
;    TITLE OF INVENTION:  PLANTS FOR ALTERED FIBER
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nicholas J. Seay, Quarles & Brady
;      STREET:  P.O. Box 2113, First Wisconsin Plaza
;      CITY:  Madison
;      STATE:  Wisconsin
;      COUNTRY:  USA
;      ZIP:  53701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Microsoft Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/298,829
;      FILING DATE:  19-OCT-1994
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/885,970
;      FILING DATE:  18-MAY-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/617,239
;      FILING DATE:  21-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/253,243
;      FILING DATE:  04-OCT-1988
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seay, Nicholas J.
;      REGISTRATION NUMBER:  27,386
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (608) 283-2478
;      TELEFAX:  (608) 251-5139
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  913 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Gossypium hirsutum
;      STRAIN:  Coker 312
;      DEVELOPMENTAL STAGE:  15 day old fiber cells
;      TISSUE TYPE:  fiber cells
```

```
;    IMMEDIATE SOURCE:
;      LIBRARY:  CKFB15A1
;      CLONE:  H6
US-08-298-829-3


  Query Match             3.8%;  Score 36.6;  DB 1;  Length 913;
  Best Local Similarity   58.9%;  Pred. No. 0.85;
  Matches   63;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy    198 tcataccaatacgacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgcc 257
          ||   ||| | |  || || | ||  || || | ||||| || || |||  | || | ||
Db    320 TCCACCCAGTTCTCCTCCTCCTGCAACTCCACCACCTGCTTCTCCTCCTCCTGCAACTCC 379

Qy    258 agccaaggcagccgcgcctcctgccgcggcgcgatcggctgagccgc 304
          | |    ||  |  | ||||||||| |  | | | | |||   || |
Db    380 ACCTCCAGCTTCTCCACCTCCTGCCACTCCTCCACCAGCTTCTCCAC 426


RESULT  15
US-08-787-335-2
; Sequence 2, Application US/08787335
; Patent No. 5981834
;  GENERAL INFORMATION:
;    APPLICANT:  John, Maliyakal E.
;    APPLICANT:  Umbeck, Paul F.
;    APPLICANT:  Brill, Winston J.
;    TITLE OF INVENTION:  GENETICALY ENGINEERED COTTON PLANTS
;    TITLE OF INVENTION:  FOR ALTERED FIBER
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Quarles and Brady
;      STREET:  P.O BOX 2113
;      STREET:  FIRST WISCONSIN PLAZA
;      CITY:  MADISON
;      STATE:  WISCONSIN
;      COUNTRY:  U.S.A.
;      ZIP:  53701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette - 3.50 inch, 800Kb storage
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Macintosh
;      SOFTWARE:  Microsoft Word 4.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/787,335
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/530,797
;      FILING DATE:
;      APPLICATION NUMBER:  US 07/253,243
;      FILING DATE:  04-OCT-88
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nicholas J. Seay
;      REGISTRATION NUMBER:  27,386
;      REFERENCE/DOCKET NUMBER:  1122990245
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  913 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  no
;    ANTI-SENSE:  no
;    ORIGINAL SOURCE:
;      ORGANISM:  Gossypium hirsutum
;      STRAIN:  Coker 312
;      DEVELOPMENTAL STAGE:  15 day old fiber cells
;      TISSUE TYPE:  fiber cells
;    IMMEDIATE SOURCE:
;      LIBRARY:  CKFB15A1
;      CLONE:  H6
US-08-787-335-2


  Query Match             3.8%;  Score 36.6;  DB 2;  Length 913;
  Best Local Similarity   58.9%;  Pred. No. 0.85;
  Matches   63;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy    198 tcataccaatacgacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgcc 257
          ||   ||| | |  || || | ||  || || | ||||| || || |||  | || | ||
Db    320 TCCACCCAGTTCTCCTCCTCCTGCAACTCCACCACCTGCTTCTCCTCCTCCTGCAACTCC 379

Qy    258 agccaaggcagccgcgcctcctgccgcggcgcgatcggctgagccgc 304
          | |    ||  |  | ||||||||| |  | | | | |.|||   || |
Db    380 ACCTCCAGCTTCTCCACCTCCTGCCACTCCTCCACCAGCTTCTCCAC 426


Search completed: September 21, 2001, 02:16:29
Job time: 28222 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: September 21, 2001, 02:15:37 ; Search time 6407.28 Seconds
(without alignments)
2250.903 Million cell updates/sec

Title: US-09-138-735-1_COPY_1232_2207
Perfect score: 976
Sequence: 1 caggtacagcgtaacggctt.....tgtgaacgaggccattatgt 976

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 14155048 seqs, 7388405095 residues

Total number of hits satisfying chosen parameters: 28310096

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*
1: /cgn1_7/ptodata/1/pna/PCTUS_COMB.seq:*
2: /cgn1_7/ptodata/1/pna/US06_COMB.seq:*
3: /cgn1_7/ptodata/1/pna/US07_COMB.seq:*
4: /cgn1_7/ptodata/1/pna/US080_COMB.seq:*
5: /cgn1_7/ptodata/1/pna/US081_COMB.seq:*
6: /cgn1_7/ptodata/1/pna/US082_COMB.seq:*
7: /cgn1_7/ptodata/1/pna/US083_COMB.seq:*
8: /cgn1_7/ptodata/1/pna/US084_COMB.seq:*
9: /cgn1_7/ptodata/1/pna/US085_COMB.seq:*
10: /cgn1_7/ptodata/1/pna/US086_COMB.seq:*
11: /cgn1_7/ptodata/1/pna/US087_COMB.seq:*
12: /cgn1_7/ptodata/1/pna/US088_COMB.seq:*
13: /cgn1_7/ptodata/1/pna/US089_COMB.seq:*
14: /cgn1_7/ptodata/1/pna/US090_COMB.seq:*
15: /cgn1_7/ptodata/1/pna/US091_COMB.seq:*
16: /cgn1_7/ptodata/1/pna/US092_COMB.seq:*
17: /cgn1_7/ptodata/1/pna/US093_COMB.seq:*
18: /cgn1_7/ptodata/1/pna/US094_COMB.seq:*
19: /cgn1_7/ptodata/1/pna/US095A_COMB.seq:*
20: /cgn1_7/ptodata/1/pna/US095B_COMB.seq:*
21: /cgn1_7/ptodata/1/pna/US095C_COMB.seq:*
22: /cgn1_7/ptodata/1/pna/US095D_COMB.seq:*
23: /cgn1_7/ptodata/1/pna/US096A_COMB.seq:*
24: /cgn1_7/ptodata/1/pna/US096B_COMB.seq:*
25: /cgn1_7/ptodata/1/pna/US096C_COMB.seq:*
26: /cgn1_7/ptodata/1/pna/US096D_COMB.seq:*
27: /cgn1_7/ptodata/1/pna/US096E_COMB.seq:*
28: /cgn1_7/ptodata/1/pna/US097A_COMB.seq:*
29: /cgn1_7/ptodata/1/pna/US097B_COMB.seq:*
30: /cgn1_7/ptodata/1/pna/US097C_COMB.seq:*
31: /cgn1_7/ptodata/1/pna/US098_COMB.seq:*
32: /cgn1_7/ptodata/1/pna/US6000_COMB.seq:*
33: /cgn1_7/ptodata/1/pna/US6001_COMB.seq:*
34: /cgn1_7/ptodata/1/pna/US6002_COMB.seq:*
35: /cgn1_7/ptodata/1/pna/US6003_COMB.seq:*
36: /cgn1_7/ptodata/1/pna/US6004_COMB.seq:*
37: /cgn1_7/ptodata/1/pna/US6005_COMB.seq:*
38: /cgn1_7/ptodata/1/pna/US6006_COMB.seq:*
39: /cgn1_7/ptodata/1/pna/US6007_COMB.seq:*
40: /cgn1_7/ptodata/1/pna/US6008_COMB.seq:*
41: /cgn1_7/ptodata/1/pna/US6009_COMB.seq:*
42: /cgn1_7/ptodata/1/pna/US6010_COMB.seq:*
43: /cgn1_7/ptodata/1/pna/US6011_COMB.seq:*
44: /cgn1_7/ptodata/1/pna/US6012_COMB.seq:*
45: /cgn1_7/ptodata/1/pna/US6013_COMB.seq:*
46: /cgn1_7/ptodata/1/pna/US6014_COMB.seq:*
47: /cgn1_7/ptodata/1/pna/US6015_COMB.seq:*
48: /cgn1_7/ptodata/1/pna/US6016_COMB.seq:*
49: /cgn1_7/ptodata/1/pna/US6017_COMB.seq:*
50: /cgn1_7/ptodata/1/pna/US6018_COMB.seq:*
51: /cgn1_7/ptodata/1/pna/US6019_COMB.seq:*
52: /cgn1_7/ptodata/1/pna/US6020_COMB.seq:*
53: /cgn1_7/ptodata/1/pna/US6021_COMB.seq:*
54: /cgn1_7/ptodata/1/pna/US6022_COMB.seq:*
55: /cgn1_7/ptodata/1/pna/US6023_COMB.seq:*
56: /cgn1_7/ptodata/1/pna/US6024_COMB.seq:*
57: /cgn1_7/ptodata/1/pna/US6025_COMB.seq:*
58: /cgn1_7/ptodata/1/pna/US6026_COMB.seq:*
59: /cgn1_7/ptodata/1/pna/US6027_COMB.seq:*
60: /cgn1_7/ptodata/1/pna/US6028_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		976	100.0	3402	13	US-08-988-242-1	Sequence 1, Appli
2		976	100.0	3402	15	US-09-138-736-1	Sequence 1, Appli
3	c	46	4.7	606	27	US-09-699-999-5885	Sequence 5885, Ap
4	c	45.8	4.7	470	27	US-09-699-999-4102	Sequence 4102, Ap
5	c	44.6	4.6	599	17	US-09-396-087-4576	Sequence 4576, Ap
6		42.6	4.4	6257	24	US-09-620-392-44819	Sequence 44819, A
7		42.6	4.4	29301	28	US-09-702-134-5721	Sequence 5721, Ap
8		42.4	4.3	600	52	US-60-209-830-45296	Sequence 45296, A
9	c	42.2	4.3	491	51	US-60-197-872-57273	Sequence 57273, A
10	c	42.2	4.3	8953	24	US-09-620-392-68153	Sequence 68153, A
11		42.2	4.3	10344	24	US-09-620-392-65686	Sequence 65686, A
12	c	42.2	4.3	10348	28	US-09-702-134-8344	Sequence 8344, Ap
13		42	4.3	7227	48	US-60-168-139-2273	Sequence 2273, Ap
14	c	42	4.3	26276	19	US-09-514-000-337	Sequence 337, App
15		42	4.3	397803	29	US-09-739-449-217	Sequence 217, App
16		41.6	4.3	600	52	US-60-207-458-67549	Sequence 67549, A
17		41.4	4.2	602	52	US-60-209-830-19436	Sequence 19436, A
18	c	41.2	4.2	322	28	US-09-716-472-1859	Sequence 1859, Ap
19	c	41.2	4.2	728	23	US-09-614-387-7832	Sequence 7832, Ap
20	c	41.2	4.2	728	28	US-09-716-472-5422	Sequence 5422, Ap
21		41.2	4.2	741	50	US-60-185-213-686	Sequence 686, App
22	c	41.2	4.2	12871	24	US-09-620-392-3460	Sequence 3460, Ap
23	c	41.2	4.2	33579	28	US-09-702-134-4946	Sequence 4946, Ap
24		40.8	4.2	399	52	US-60-207-458-74910	Sequence 74910, A
25		40.8	4.2	415	52	US-60-207-458-150303	Sequence 150303,
26		40.8	4.2	547	52	US-60-207-458-149725	Sequence 149725,
27		40.8	4.2	573	22	US-09-553-094-5427	Sequence 5427, Ap
28		40.8	4.2	635	52	US-60-209-830-60545	Sequence 60545, A
29		40.8	4.2	637	52	US-60-207-458-151348	Sequence 151348,
30		40.8	4.2	668	52	US-60-207-458-151301	Sequence 151301,
31	c	40.6	4.2	550	51	US-60-197-872-44492	Sequence 44492, A
32		40.4	4.1	470	25	US-09-654-617-111007	Sequence 111007,
33		40.4	4.1	470	27	US-09-684-016-111007	Sequence 111007,
34	c	40.4	4.1	477	15	US-09-135-994-1	Sequence 1, Appli
35		40.4	4.1	569	53	US-60-211-750-30151	Sequence 30151, A
36		40.4	4.1	5537	1	PCT-US01-04098A-2481	Sequence 2481, Ap
37		40.4	4.1	5537	18	US-09-496-914A-6707	Sequence 6707, Ap
38		40.4	4.1	5537	22	US-09-560-875A-6707	Sequence 6707, Ap
39	c	40.4	4.1	6248	1	PCT-US01-04098A-513	Sequence 513, App
40		40.4	4.1	4403765	15	US-09-103-840A-2	Sequence 2, Appli
41		40.4	4.1	4411529	15	US-09-103-840A-1	Sequence 1, Appli
42	c	40.2	4.1	811	25	US-09-654-617-392521	Sequence 392521,
43	c	40.2	4.1	811	27	US-09-684-016-392521	Sequence 392521,
44	c	40	4.1	6867	24	US-09-620-392-66151	Sequence 66151, A
45	c	40	4.1	7443	24	US-09-620-392-55872	Sequence 55872, A

ALIGNMENTS

```
RESULT   1
US-08-968-242-1
; Sequence 1, Application US/08988242
;  GENERAL INFORMATION:
;    APPLICANT:  PARANHOS-BACCALA, GLAUCIA
;    APPLICANT:  LESENECHAL, MYLENE
;    APPLICANT:  JOLIVET, MICHEL
;    APPLICANT:  MANDRAND, BERNARD
;    TITLE OF INVENTION:  NEW TRYPANOSOMA CRUZI ANTIGEN, GENE
;    TITLE OF INVENTION:  ENCODING THEREFOR, AND METHODS OF DETECTING AND TREATING
;    TITLE OF INVENTION:  CHAGAS DISEASE
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OLIFF & BERRIDGE, PLC
;      STREET:  P.O. BOX 19928
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22320
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/988,242
;      FILING DATE:  10-DEC-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berridge, William P.
;      REGISTRATION NUMBER:  30,024
;      REFERENCE/DOCKET NUMBER:  WPB 36400A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-836-6400
;      TELEFAX:  703-836-2787
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3402 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-988-242-1


  Query Match             100.0%;  Score 976;  DB 13;  Length 3402;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-221;
  Matches  976;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 caggtacagcgtaacggcttttgcttcaatcgtacagccgacggtagctgcgtcctggct 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1232 CAGGTACAGCGTAACGGCTTTTGCTTCAATCGTACAGCCGACGGTAGCTGCGTCCTGGCT 1291

Qy     61 gacatgtcgattcgattgacgatcttccatctccggtcctcccgcagggaagaacagcag 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1292 GACATGTCGATTCGATTGACGATCTTCCATCTCCGGTCCTCCCGCAGGGAAGAACAGCAG 1351

Qy    121 ccaggccaaaaaacatcggtagtggcgacggcgaaaccggggtgtgtgtcctcgggcact 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1352 CCAGGCCAAAAAACATCGGTAGTGGCGACGGCGAAACCGGGGTGTGTGTCCTCGGGCACT 1411

Qy    181 gacgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatcccctgcatca 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1412 GACGCGGCGAGTAGCAGTCATACCAATACGACTTCTGCCGCTGCTGCATCCCCTGCATCA 1471

Qy    241 ccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggctgag 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1472 CCCCCTGTTTCAGCGCCAGCCAAGGCAGCCGCGCCTCCTGCCGCGGCGCGATCGGCTGAG 1531

Qy    301 ccgcacgtggggagcaagatcattgctaatctagtgaatcagctggggattaatgtcacc 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1532 CCGCACGTGGGGAGCAAGATCATTGCTAATCTAGTGAATCAGCTGGGGATTAATGTCACC 1591

Qy    361 caaaggagcgtcgtcagcactggagcgccggccacgacgaggtctacggcggtgacgtcc 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1592 CAAAGGAGCGTCGTCAGCACTGGAGCGCCGGCCACGACGAGGTCTACGGCGGTGACGTCC 1651

Qy    421 acgactaccgccccgcagcgaacaagtccatacgggcacaatggccgacctgtgacggct 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1652 ACGACTACCGCCCCGCAGCGAACAAGTCCATACGGGCACAATGGCCGACCTGTGACGGCT 1711

Qy    481 ggattggtggcagctaatagtggtgccagcgcggcctcgtctcccacagccgcggcgaaa 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1712 GGATTGGTGGCAGCTAATAGTGGTGCCAGCGCGGCCTCGTCTCCCACAGCCGCGGCGAAA 1771

Qy    541 ccaacaggagaagaaaaggcctccgcggcatgtgaaacgagctccgtggcgataaatgcg 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1772 CCAACAGGAGAAGAAAAGGCCTCCGCGGCATGTGAAACGAGCTCCGTGGCGATAAATGCG 1831

Qy    601 acacgcccggcgcttcacaacgcctctctcccgcaggcgccaacggatggcgttttggcg 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1832 ACACGCCCGGCGCTTCACAACGCCTCTCTCCCGCAGGCGCCAACGGATGGCGTTTTGGCG 1891

Qy    661 gcagcagtataccagtcggagggcgaggttcatcagtcgctggagcggctggagtccgtc 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1892 GCAGCAGTATACCAGTCGGAGGGCGAGGTTCATCAGTCGCTGGAGCGGCTGGAGTCCGTC 1951

Qy    721 ataaccaacacgtctcgggttctgaagttgctccctgacaccattcgaagagaccatgaa 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1952 ATAACCAACACGTCTCGGGTTCTGAAGTTGCTCCCTGACACCATTCGAAGAGACCATGAA 2011

Qy    781 caacttctgaatctgggtttagaggcacagatgacagagctgcagcagagccgtccaaca 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2012 CAACTTCTGAATCTGGGTTTAGAGGCACAGATGACAGAGCTGCAGCAGAGCCGTCCAACA 2071

Qy    841 ccgcaaacacagccgagagacacaagctccgcgaaatcatccgtgtttgagacgtacacc 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2072 CCGCAAACACAGCCGAGAGACACAAGCTCCGCGAAATCATCCGTGTTTGAGACGTACACC 2131

Qy    901 cttgttctcattgcggattccctctctcgcaacatcacgaagggggtgaagcgtggtgtg 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2132 CTTGTTCTCATTGCGGATTCCCTCTCTCGCAACATCACGAAGGGGGTGAAGCGTGGTGTG 2191

Qy    961 aacgaggccattatgt 976
          ||||||||||||||||
Db   2192 AACGAGGCCATTATGT 2207


RESULT   2
US-09-138-736-1
; Sequence 1, Application US/09138736
;  GENERAL INFORMATION:
;    APPLICANT:  PARANHOS-BACCALA, Glaucia
;    APPLICANT:  LESENECHAL, Mylene
;    APPLICANT:  JOLIVET, Michel
;    TITLE OF INVENTION:  NEW TRYPANOSOMA CRUZI ANTIGEN, AND GENE
;    TITLE OF INVENTION:  ENCODING THE LATTER; THEIR APPLICATION TO THE DETECTION OF
;    TITLE OF INVENTION:  CHAGAS DISEASE
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Oliff & Berridge
;      STREET:  700 South Washington Street, Suite 300
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22314
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
```

```
;       APPLICATION NUMBER:  US/09/138,736
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/480,917
;       FILING DATE:  07-JUN-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Berridge, William P.
;       REGISTRATION NUMBER:  30,024
;       REFERENCE/DOCKET NUMBER:  WPB 36400
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  703-836-6400
;       TELEFAX:  703-836-2787
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3402 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-09-138-736-1


  Query Match           100.0%;  Score 976;  DB 15;  Length 3402;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-221;
  Matches  976;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 caggtacagcgtaacggctttttgcttcaatcgtacagccgacggtagctgcgtcctggct 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1232 CAGGTACAGCGTAACGGCTTTTGCTTCAATCGTACAGCCGACGGTAGCTGCGTCCTGGCT 1291

Qy     61 gacatgtcgattcgattgacgatcttccatctccggtcctcccgcagggaagaacagcag 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1292 GACATGTCGATTCGATTGACGATCTTCCATCTCCGGTCCTCCCGCAGGGAAGAACAGCAG 1351

Qy    121 ccaggccaaaaaacatcggtagtggcgacggcgaaaccggggtgtgtgtcctcgggcact 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1352 CCAGGCCAAAAAACATCGGTAGTGGCGACGGCGAAACCGGGGTGTGTGTCCTCGGGCACT 1411

Qy    181 gacgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatcccctgcatca 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1412 GACGCGGCGAGTAGCAGTCATACCAATACGACTTCTGCCGCTGCTGCATCCCCTGCATCA 1471

Qy    241 ccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggctgag 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1472 CCCCCTGTTTCAGCGCCAGCCAAGGCAGCCGCGCCTCCTGCCGCGGCGCGATCGGCTGAG 1531

Qy    301 ccgcacgtggggagcaagatcattgctaatctagtgaatcagctggggattaatgtcacc 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1532 CCGCACGTGGGGAGCAAGATCATTGCTAATCTAGTGAATCAGCTGGGGATTAATGTCACC 1591

Qy    361 caaaggagcgtcgtcagcactggagcgccggccacgacgaggtctacggcggtgacgtcc 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1592 CAAAGGAGCGTCGTCAGCACTGGAGCGCCGGCCACGACGAGGTCTACGGCGGTGACGTCC 1651

Qy    421 acgactaccgccccgcagcgaacaagtccatacgggcacaatggccgacctgtgacggct 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1652 ACGACTACCGCCCCGCAGCGAACAAGTCCATACGGGCACAATGGCCGACCTGTGACGGCT 1711

Qy    481 ggattggtggcagctaatagtggtgccagcgcggcctcgtctcccacagccgcggcgaaa 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1712 GGATTGGTGGCAGCTAATAGTGGTGCCAGCGCGGCCTCGTCTCCCACAGCCGCGGCGAAA 1771

Qy    541 ccaacaggagaagaaaaggcctccgcggcatgtgaaacgagctccgtggcgataaatgcg 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1772 CCAACAGGAGAAGAAAAGGCCTCCGCGGCATGTGAAACGAGCTCCGTGGCGATAAATGCG 1831

Qy    601 acacgcccggcgcttcacaacgcctctctcccgcaggcgccaacggatggcgttttggcg 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1832 ACACGCCCGGCGCTTCACAACGCCTCTCTCCCGCAGGCGCCAACGGATGGCGTTTTGGCG 1891

Qy    661 gcagcagtataccagtcggagggcgaggttcatcagtcgctggagcggctggagtccgtc 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1892 GCAGCAGTATACCAGTCGGAGGGCGAGGTTCATCAGTCGCTGGAGCGGCTGGAGTCCGTC 1951

Qy    721 ataaccaacacgtctcgggttctgaagttgctcccctgacaccattcgaagagaccatgaa 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1952 ATAACCAACACGTCTCGGGTTCTGAAGTTGCTCCCTGACACCATTCGAAGAGACCATGAA 2011

Qy    781 caacttctgaatctgggtttagaggcacagatgacagagctgcagcagagccgtccaaca 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2012 CAACTTCTGAATCTGGGTTTAGAGGCACAGATGACAGAGCTGCAGCAGAGCCGTCCAACA 2071

Qy    841 ccgcaaacacagccgagagacacaagctccgcgaaatcatccgtgtttgagacgtacacc 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2072 CCGCAAACACAGCCGAGAGACACAAGCTCCGCGAAATCATCCGTGTTTGAGACGTACACC 2131

Qy    901 cttgttctcattgcggattccctctctcgcaacatcacgaaggggtgaagcgtggtgtg 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2132 CTTGTTCTCATTGCGGATTCCCTCTCTCGCAACATCACGAAGGGGGTGAAGCGTGGTGTG 2191

Qy    961 aacgaggccattatgt 976
          ||||||||||||||||
Db   2192 AACGAGGCCATTATGT 2207


RESULT   3
US-09-699-999-5885/c
; Sequence 5885, Application US/09699999
; GENERAL INFORMATION:
;  APPLICANT:  Shyjan, Andrew W.
;  APPLICANT:  MacBeth, Kyle J.
;  APPLICANT:  Vasicek, Thomas
;  APPLICANT:  Wang, Youzhan
;  TITLE OF INVENTION:  NOVEL NUCLEIC ACID MOLECULES AND USES
;  TITLE OF INVENTION:  THEREFOR
;  FILE REFERENCE:  1600.2032-001
;  CURRENT APPLICATION NUMBER: US/09/699,999
;  CURRENT FILING DATE:  2000-10-30
;  PRIOR APPLICATION NUMBER:  60/162,361
;  PRIOR FILING DATE:  1999-10-29
;  NUMBER OF SEQ ID NOS:  7488
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 5885
;   LENGTH: 606
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(606)
;   OTHER INFORMATION: n = A,T,C or G
US-09-699-999-5885


  Query Match             4.7%;  Score 46;  DB 27;  Length 606;
  Best Local Similarity   46.5%;  Pred. No. 1.1;
  Matches  148;  Conservative    0;  Mismatches  170;  Indels    0;  Gaps    0;

Qy    106 agggaagaacagcagccaggccaaaaaacatcggtagtggcgacggcgaaaccggggtgt 165
          ||||  ||  | |     ||| . || ||  |||  | |||| ||| |      ||| |
Db    324 AGGGCTGAGGACCCCGGGGGCATGAACACGCCGGAGGGGGCGGCGGTGGGGGGCGGGCGG 265

Qy    166 gtgtcctcgggcactgacgcggcgagtagcagtcataccaatacgacttctgccgctgct 225
            |  | ||  ||      |||     |       |   ||    || || |||| |  |||
Db    264 CTTCCTTCAGGAGGGCGCGCCCGCCGAGCGCCGCTCTCCGCGGCGGCTGCTGCTGTGGCT 205

Qy    226 gcatcccctgcatcacccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcg 285
          |   |  ||||  |  | |    | ||| | | ||    |  || | | || |||| ||
Db    204 GTCGCTGCTGCTGCCGCTCTAACTGCAGTGGCGGCTGCTGTGGCTGTGGCTGCTGCTGCT 145

Qy    286 gcgcgatcggctgagccgcacgtggggagcaagatcattgctaatctagtgaatcagctg 345
            | ||  | || |   ||  | | |  | |  | |     || |||   | || |
Db    144 CCCCGGGCCGCAGGCGCGTCCGCGTCGTGGCCGCTGCTGGGATCGCTGGTGCCGCCGCCG 85

Qy    346 gggattaatgtcacccaaaggagcgtcgtcagcactggagcgccggccacgacgaggtct 405
            |      | | |||    |   || |   || | |   | | |||| |   | ||
```

```
Db      84 CCGCCGCCAGGCGCCCGCTCGCATGTGGCTCGCGCCGCGTCTCAGGCCGCCCGGCGGATC 25

Qy     406 acggcggtgacgtccacg 423
           ||||||   |||   ||
Db      24 CCGGCGGACGCGTGGGCG 7


RESULT   4
US-09-699-999-4102/c
; Sequence 4102, Application US/09699999
; GENERAL INFORMATION:
;  APPLICANT:  Shyjan, Andrew W.
;  APPLICANT:  MacBeth, Kyle J.
;  APPLICANT:  Vasicek, Thomas
;  APPLICANT:  Wang, Youzhan
;  TITLE OF INVENTION:  NOVEL NUCLEIC ACID MOLECULES AND USES
;  TITLE OF INVENTION:  THEREFOR
;  FILE REFERENCE:  1600.2032-001
;  CURRENT APPLICATION NUMBER: US/09/699,999
;  CURRENT FILING DATE:  2000-10-30
;  PRIOR APPLICATION NUMBER:  60/162,361
;  PRIOR FILING DATE:  1999-10-29
;  NUMBER OF SEQ ID NOS:  7488
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 4102
;   LENGTH: 470
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(470)
;   OTHER INFORMATION: n = A,T,C or G
US-09-699-999-4102


  Query Match             4.7%;  Score 45.8;  DB 27;  Length 470;
  Best Local Similarity   46.6%;  Pred. No. 1.2;
  Matches  146;  Conservative    0;  Mismatches  167;  Indels    0;  Gaps    0;

Qy     106 agggaagaacagcagccaggccaaaaaaacatcggtagtggcgacggcgaaaccggggtgt 165
           ||||  ||  | |     |||   || ||  |||  | |||| ||| |       ||| |
Db     319 AGGGCTGAGGACCCCGGGGGCATGAACACGCCGGAGGGGGCGGCGGTGGGGGGCGGGCGG 260

Qy     166 gtgtcctcgggcactgacgcggcgagtagcagtcataccaatacgacttctgccgctgct 225
           |  | || ||      |||     |        |   ||   || || |||| |  |||
Db     259 CTTCCTTCAGGAGGGCGCGCCCGCCGAGCGCCGCTCTCCGCGGCGGCTGCTGCTGTGGCT 200

Qy     226 gcatcccctgcatcacccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcg 285
           |   |  ||||  |  | |    | ||| | | ||    |  || | | || |||| ||
Db     199 GTCGCTGCTGCTGCCGCTCTAACTGCAGTGGCGGCTGCTGTGGCTGTGGCTGCTGCTGCT 140

Qy     286 gcgcgatcggctgagccgcacgtggggagcaagatcattgctaatctagtgaatcagctg 345
           | ||  | || |   ||  || | | |   | |  | |     || |||   | || |
Db     139 CCCCGGGCCGCAGGCGCGTCCGCGTCGTGGCCGCTGCTGGGATCGCTGGTGCCGCCGCCG 80

Qy     346 gggattaatgtcacccaaaggagcgtcgtcagcactggagcgccggccacgacgaggtct 405
           |      | | |||    |   || |   || | |   | | |||| |   | ||
Db      79 CCGCCGCCAGGCGCCCGCTCGCATGTGGCTCGCGCCGCGTCTCAGGCCGCCCGGCGGATC 20

Qy     406 acggcggtgacgt 418
           ||||||   |||
Db      19 CCGGCGGACGCGT 7


RESULT   5
US-09-396-087-4576/c
; Sequence 4576, Application US/09396087
; GENERAL INFORMATION:
;  APPLICANT: Gearing, David P.
;  APPLICANT: Holtzman, Douglas A.
;  TITLE OF INVENTION: NUCLEIC ACID MOLECULES DERIVED FROM A
;  TITLE OF INVENTION: PULMONARY ARTERY SMOOTH MUSCLE LIBRARY
;  FILE REFERENCE: MLN98-39pM
;  CURRENT APPLICATION NUMBER: US/09/396,087
;  CURRENT FILING DATE: 1999-09-14
;  EARLIER APPLICATION NUMBER: 60/100,260
;  EARLIER FILING DATE: 1998-09-14
;  EARLIER APPLICATION NUMBER: 60/107,226
;  EARLIER FILING DATE: 1998-11-05
;  EARLIER APPLICATION NUMBER: 60/131,810
;  EARLIER FILING DATE: 1999-04-30
;  NUMBER OF SEQ ID NOS: 5220
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4576
;   LENGTH: 599
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(599)
;   OTHER INFORMATION: n = A,T,C or G
US-09-396-087-4576


  Query Match             4.6%;  Score 44.6;  DB 17;  Length 599;
  Best Local Similarity   46.6%;  Pred. No. 2.3;
  Matches  143;  Conservative    0;  Mismatches  164;  Indels    0;  Gaps    0;

Qy     106 agggaagaacagcagccaggccaaaaaaacatcggtagtggcgacggcgaaaccggggtgt 165
           ||||  ||  | |     |||   || ||  |||  | |||| ||| |       ||| |
Db     316 AGGGCTGAGGACCCCGGGGGCATGAACACGCCGGAGGGGGCGGCGGTGGGGGGCGGGCGG 257

Qy     166 gtgtcctcgggcactgacgcggcgagtagcagtcataccaatacgacttctgccgctgct 225
           |  | || ||      |||     |        |   ||   || || |||| |  |||
Db     256 CTTCCTTCAGGAGGGCGCGCCCGCCGAGCGCCGCTCTCCGCGGCGGCTGCTGCTGTGGCT 197

Qy     226 gcatcccctgcatcacccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcg 285
           |   |  ||||  |  | |    | ||| | | ||    |  || | | || |||| ||
Db     196 GTCGCTGCTGCTGCCGCTCTAACTGCAGTGGCGGCTGCTGTGGCTGTGGCTGCTGCTGCT 137

Qy     286 gcgcgatcggctgagccgcacgtggggagcaagatcattgctaatctagtgaatcagctg 345
           | ||  | || |   ||  || | | |   | |  | |     || |||   | || |
Db     136 CCCCGGGCCGCAGGCGCGTCCGCGTCGTGGCCGCTGCTGGGATCGCTGGTGCCGCCGCCG 77

Qy     346 gggattaatgtcacccaaaggagcgtcgtcagcactggagcgccggccacgacgaggtct 405
           |      | | |||    |   || |   || | |   | | |||| |   | ||
Db      76 CCGCCGCCAGGCGCCCGCTCGCATGTGGCTCGCGCCGCGTCTCAGGCCGCCCGGCGGATC 17

Qy     406 acggcgg 412
           |||||||
Db      16 CCGGCGG 10


RESULT   6
US-09-620-392-44819
; Sequence 44819, Application US/09620392
; GENERAL INFORMATION:
;  APPLICANT:       Boukharov, Andrey A.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Liu, Jingdong
;  APPLICANT:  McIninch, James
;  TITLE OF INVENTION:       Plant Genome Sequence And Uses Thereof
;  FILE REFERENCE:       38-21(51237)E
;  CURRENT APPLICATION NUMBER: US/09/620,392
;  CURRENT FILING DATE:  2000-07-19
;  NUMBER OF SEQ ID NOS:       69652
; SEQ ID NO 44819
;   LENGTH: 6257
;   TYPE: DNA
;   ORGANISM: Oryza sativa
US-09-620-392-44819


  Query Match             4.4%;  Score 42.6;  DB 24;  Length 6257;
  Best Local Similarity   48.2%;  Pred. No. 9.5;
  Matches  120;  Conservative    0;  Mismatches  129;  Indels    0;  Gaps    0;
```

```
Qy    161 ggtgtgtgtcctcgggcactgacgcggcgagtagcagtcataccaatacgacttctgccg 220
          | |  | | || |||   | |  ||||      ||        ||   |||| | || |
Db   2350 gttcggcggccccggcggccgcggcggacccgcgctcggccctcacggcgacgacggcgg 2409

Qy    221 ctgctgcatcccctgcatcacccccctgtttcagcgccagccaaggcagccgcgcctcctg 280
          |  | ||  | || ||  |   | | |  || | | |  |  |||| || ||| |  | |
Db   2410 caccggcggcaccggcggctttctcggcgtcggaggcgacggaggccgcggcggcgacgg 2469

Qy    281 ccgcggcgcgatcggctgagccgcacgtggggagcaagatcattgctaatctagtgaatc 340
          | ||   |    | || ||| |    | || ||   || |||  ||  | || |  ||
Db   2470 cggcccaggaggcagcggaggccgcggcggcgacggaggtcaaggcggacctggccaacg 2529

Qy    341 agctggggattaatgtcacccaaaggagcgtcgtcagcactggagcgccggccacgacga 400
           |  |||||     | | |     ||   |  | |    | ||  || |||| |||  |
Db   2530 gggcggggacggcgggcccggcggggcaggcgggcccggcgggcacggcggcgacggtgg 2589

Qy    401 ggtctacgg 409
          || |  |||
Db   2590 ggacggcgg 2598


RESULT   7
US-09-702-134-5721
; Sequence 5721, Application US/09702134
; GENERAL INFORMATION:
;  APPLICANT:       Boukharov, Andrey A.
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Liu, Jingdong
;  APPLICANT:  McIninch, James
;  APPLICANT:  Wu, Wei
;  TITLE OF INVENTION:       Plant Genome Sequence and Uses Thereof
;  FILE REFERENCE:       38-21(51237)F
;  CURRENT APPLICATION NUMBER: US/09/702,134
;  CURRENT FILING DATE:  2000-10-31
;  NUMBER OF SEQ ID NOS:       52202
; SEQ ID NO 5721
;   LENGTH: 29301
;   TYPE: DNA
;   ORGANISM: Oryza sativa
US-09-702-134-5721


  Query Match             4.4%;  Score 42.6;  DB 28;  Length 29301;
  Best Local Similarity   48.2%;  Pred. No. 12;
  Matches  120;  Conservative    0;  Mismatches  129;  Indels    0;  Gaps    0;

Qy    161 ggtgtgtgtcctcgggcactgacgcggcgagtagcagtcataccaatacgacttctgccg 220
          | |  | | || |||   | |  ||||      ||        ||   |||| | || |
Db   2414 gttcggcggccccggcggccgcggcggacccgcgctcggccctcacggcgacgacggcgg 2473

Qy    221 ctgctgcatcccctgcatcacccccctgtttcagcgccagccaaggcagccgcgcctcctg 280
          |  | ||  | || ||  |   | | |  || | | |  |  |||| || ||| |  | |
Db   2474 caccggcggcaccggcggctttctcggcgtcggaggcgacggaggccgcggcggcgacgg 2533

Qy    281 ccgcggcgcgatcggctgagccgcacgtggggagcaagatcattgctaatctagtgaatc 340
          | ||   |    | || ||| |    | || ||   || |||  ||  | || |  ||
Db   2534 cggcccaggaggcagcggaggccgcggcggcgacggaggtcaaggcggacctggccaacg 2593

Qy    341 agctggggattaatgtcacccaaaggagcgtcgtcagcactggagcgccggccacgacga 400
           |  |||||     | | |     ||   |  | |    | ||  || |||| |||  |
Db   2594 gggcggggacggcgggcccggcggggcaggcgggcccggcgggcacggcggcgacggtgg 2653

Qy    401 ggtctacgg 409
          || |  |||
Db   2654 ggacggcgg 2662


RESULT   8
US-60-209-830-45296
; Sequence 45296, Application US/60209830
; GENERAL INFORMATION:
;  APPLICANT:       Andersen, Scott E.
;  APPLICANT:  Castiglioni, Paolo
;  APPLICANT:  Hardeman, Kristine J.
;  APPLICANT:  Kerk, Nancy M.
;  APPLICANT:  La Rosa, Thomas J.
;  TITLE OF INVENTION:       NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WI
;  TITLE OF INVENTION:  PLANTS
;  FILE REFERENCE:       38-21(51934)A
;  CURRENT APPLICATION NUMBER: US/60/209,830
;  CURRENT FILING DATE:  2000-06-06
;  NUMBER OF SEQ ID NOS:       62628
; SEQ ID NO 45296
;   LENGTH: 600
;   TYPE: DNA
;   ORGANISM: Zea mays
;   OTHER INFORMATION: Clone ID: LIB3596-038-Q6-K6-G7
US-60-209-830-45296


  Query Match             4.3%;  Score 42.4;  DB 52;  Length 600;
  Best Local Similarity   60.3%;  Pred. No. 7.7;
  Matches   70;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy    178 actgacgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatcccctgca 237
          || | | | |||||| || |||  |||  | | ||  | ||| || | ||  | ||| |
Db    399 acagccccagcgagtcgccgtcgcacccctcctacggccgccactcccgctcctcctccg 458

Qy    238 tcacccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgatc 293
          ||| || ||  | | ||| |  ||| ||  || ||| |  | | |||   | | ||
Db    459 tcagccgcttctccggcggcctccacgggggcggcggcggcgggcgcaaggtgctc 514


RESULT   9
US-60-197-872-57273/c
; Sequence 57273, Application US/60197872
; GENERAL INFORMATION:
;  APPLICANT:       Bougri, Olegs
;  APPLICANT:  Byrum, Joseph R.
;  APPLICANT:  De La Pena, Robert C.
;  APPLICANT:  La Rosa, Thomas J.
;  APPLICANT:  Shukla, Hridayabhiranjan
;  TITLE OF INVENTION:       Nucleic acid Molecules and Other molecules associated wi
;  TITLE OF INVENTION:  plants
;  FILE REFERENCE:       38-21(51892)A
;  CURRENT APPLICATION NUMBER: US/60/197,872
;  CURRENT FILING DATE:  2000-04-19
;  NUMBER OF SEQ ID NOS:       76255
; SEQ ID NO 57273
;   LENGTH: 491
;   TYPE: DNA
;   ORGANISM: Oryza sativa nipponbare
;   OTHER INFORMATION: Clone ID: jC-osflLIB3479034a01b1
US-60-197-872-57273


  Query Match             4.3%;  Score 42.2;  DB 51;  Length 491;
  Best Local Similarity   53.3%;  Pred. No. 8.4;
  Matches   89;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy    142 gtggcgacggcgaaaccggggtgtgtgtcctcgggcactgacgcggcgagtagcagtcat 201
          | |||  || |     |||  ||||  ||| | || ||||  |   || ||  |||||
Db    371 GCGGCAGCGTCAGCGGCGGCTTGTGGTTCCGCAGGTACTGCAGGAACGTGTTCCAGTCCC 312

Qy    202 accaatacgacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgccagcc 261
            ||  | |  |  ||    |     || ||||  |  | ||  |   | | |||||||||
Db    311 GCCGCTTCTGCGACTCGTACCTGCTCAGCCCCGCCGCCGCCGGCGCATGCTGCGCCAGCG 252

Qy    262 aaggcagccgcgcctcctgccgcggcgcgatcggctgagccgcacgt 308
            || || |||| || |||  | ||     || ||| | |  | ||
Db    251 GCTGCGGCTGCGCTTCTTGCACCTGCTGCTGCGCCTGCGACTGATGT 205


RESULT  10
```

```
US-09-620-392-68153/c
; Sequence 68153, Application US/09620392
; GENERAL INFORMATION:
;  APPLICANT:       Boukharov, Andrey A.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Liu, Jingdong
;  APPLICANT:  McIninch, James
;  TITLE OF INVENTION:       Plant Genome Sequence And Uses Thereof
;  FILE REFERENCE:       38-21(51237)E
;  CURRENT APPLICATION NUMBER: US/09/620,392
;  CURRENT FILING DATE:  2000-07-19
;  NUMBER OF SEQ ID NOS:       69652
; SEQ ID NO 68153
;   LENGTH: 8953
;   TYPE: DNA
;   ORGANISM: Oryza sativa
US-09-620-392-68153


  Query Match             4.3%;  Score 42.2;  DB 24;  Length 8953;
  Best Local Similarity   51.9%;  Pred. No. 12;
  Matches   95;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy    133 acatcggtagtggcgacggcgaaaccggggtgtgtgtcctcgggcactgacgcggcgagt 192
          ||| | |     ||| |  |   |||| | ||  | || |||   |    ||||| ||
Db   1548 ACACCCGCGCCTGCGCCTCCAGCACCGTCGCGTTCGGCCGCGGCATCCCCGGCGGCCGGT 1489

Qy    193 agcagtcataccaatacgacttctgccgctgctgcatcccctgcatcacccccctgtttca 252
          |  ||||  ||     ||  ||| ||| |  |  |  | || ||| | || ||     |
Db   1488 ACGAGTCCGACACCGCCGGGTTCCGCCTCACCCCCTCCTCCAGCACCGCCACCCCCGCCT 1429

Qy    253 gcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggctgagccgcacgtgggg 312
          | ||| ||| | |  |  |||| ||  || || ||| | |  |    || | ||  |
Db   1428 CCTCCATCCACGACGCCTCCGCCCCCGCCCCCGCCGCCACCACCGCCACCCCGCGCCGCC 1369

Qy    313 agc 315
          | |
Db   1368 ATC 1366


RESULT  11
US-09-620-392-65686
; Sequence 65686, Application US/09620392
; GENERAL INFORMATION:
;  APPLICANT:       Boukharov, Andrey A.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Liu, Jingdong
;  APPLICANT:  McIninch, James
;  TITLE OF INVENTION:       Plant Genome Sequence And Uses Thereof
;  FILE REFERENCE:       38-21(51237)E
;  CURRENT APPLICATION NUMBER: US/09/620,392
;  CURRENT FILING DATE:  2000-07-19
;  NUMBER OF SEQ ID NOS:       69652
; SEQ ID NO 65686
;   LENGTH: 10344
;   TYPE: DNA
;   ORGANISM: Oryza sativa
US-09-620-392-65686


  Query Match             4.3%;  Score 42.2;  DB 24;  Length 10344;
  Best Local Similarity   51.9%;  Pred. No. 13;
  Matches   95;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy    133 acatcggtagtggcgacggcgaaaccggggtgtgtgtcctcgggcactgacgcggcgagt 192
          ||| | |     ||| |  |   |||| | ||  | || |||   |    ||||| ||
Db   8556 acacccgcgcctgcgcctccagcaccgtcgcgttcggccgcggcatccccggcggccggt 8615

Qy    193 agcagtcataccaatacgacttctgccgctgctgcatcccctgcatcacccccctgtttca 252
          |  ||||  ||     ||  ||| ||| |  |  |  | || ||| | || ||     |
Db   8616 acgagtccgacaccgccgggttccgcctcacccccctcctccagcaccgccacccccgcct 8675

Qy    253 gcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggctgagccgcacgtgggg 312
          | ||| ||| | |  |  |||| ||  || || ||| | |  |    || | ||  |
Db   8676 cctccatccacgacgcctccgcccccgcccccgccgccaccaccgccaccccgcgccgcc 8735

Qy    313 agc 315
          | |
Db   8736 atc 8738


RESULT  12
US-09-702-134-8344/c
; Sequence 8344, Application US/09702134
; GENERAL INFORMATION:
;  APPLICANT:       Boukharov, Andrey A.
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Liu, Jingdong
;  APPLICANT:  McIninch, James
;  APPLICANT:  Wu, Wei
;  TITLE OF INVENTION:       Plant Genome Sequence and Uses Thereof
;  FILE REFERENCE:       38-21(51237)F
;  CURRENT APPLICATION NUMBER: US/09/702,134
;  CURRENT FILING DATE:  2000-10-31
;  NUMBER OF SEQ ID NOS:       52202
; SEQ ID NO 8344
;   LENGTH: 10348
;   TYPE: DNA
;   ORGANISM: Oryza sativa
US-09-702-134-8344


  Query Match             4.3%;  Score 42.2;  DB 28;  Length 10348;
  Best Local Similarity   51.9%;  Pred. No. 13;
  Matches   95;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy    133 acatcggtagtggcgacggcgaaaccggggtgtgtgtcctcgggcactgacgcggcgagt 192
          ||| | |     ||| |  |   |||| | ||  | || |||   |    ||||| ||
Db   1793 ACACCCGCGCCTGCGCCTCCAGCACCGTCGCGTTCGGCCGCGGCATCCCCGGCGGCCGGT 1734

Qy    193 agcagtcataccaatacgacttctgccgctgctgcatcccctgcatcacccccctgtttca 252
          |  ||||  ||     ||  ||| ||| |  |  |  | || ||| | || ||     |
Db   1733 ACGAGTCCGACACCGCCGGGTTCCGCCTCACCCCCTCCTCCAGCACCGCCACCCCCGCCT 1674

Qy    253 gcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggctgagccgcacgtgggg 312
          | ||| ||| | |  |  |||| ||  || || ||| | |  |    || | ||  |
Db   1673 CCTCCATCCACGACGCCTCCGCCCCCGCCCCCGCCGCCACCACCGCCACCCCGCGCCGCC 1614

Qy    313 agc 315
          | |
Db   1613 ATC 1611


RESULT  13
US-60-168-139-2273
; Sequence 2273, Application US/60168139
; GENERAL INFORMATION:
;  APPLICANT: Hinkle, Gregory J.
;  APPLICANT: Slater, Steven C.
;  TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequence And Uses Thereof
;  FILE REFERENCE: 38-21(15490)A
;  CURRENT APPLICATION NUMBER: US/60/168,139
;  CURRENT FILING DATE: 1999-12-01
;  NUMBER OF SEQ ID NOS: 3432
; SEQ ID NO 2273
;   LENGTH: 7227
;   TYPE: DNA
;   ORGANISM: Agrobacterium tumefaciens
;   FEATURE:
;   OTHER INFORMATION: unsure at all n locations
US-60-168-139-2273


  Query Match             4.3%;  Score 42;  DB 48;  Length 7227;
  Best Local Similarity   60.5%;  Pred. No. 13;
```

```
  Matches   69;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy    183 cgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatcccctgcatcacc 242
          | | || | || |  ||  |     ||  || | || || || ||| ||||| || ||||
Db   1696 ccccgccaatatcgttcccaaactgacacctcccgcggcggcggcagccccctccagcacc 1755

Qy    243 ccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggc 296
              ||| | || || ||   || |   || ||||| |||||||| | | | ||
Db   1756 ggaagttgctgcaccggcagcggtacttgctcctccagccgcggctccacccgc 1809


RESULT  14
US-09-514-000-337/c
; Sequence 337, Application US/09514000
; GENERAL INFORMATION:
;  APPLICANT: Hinkle, Gregory J.
;  APPLICANT: Slater, Steven C.
;  TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
;  FILE REFERENCE: 38-10(15490)B
;  CURRENT APPLICATION NUMBER: US/09/514,000
;  CURRENT FILING DATE: 2000-02-23
;  NUMBER OF SEQ ID NOS: 15034
; SEQ ID NO 337
;   LENGTH: 26276
;   TYPE: DNA
;   ORGANISM: Agrobacterium tumefaciens
US-09-514-000-337


  Query Match             4.3%;  Score 42;  DB 19;  Length 26276;
  Best Local Similarity   60.5%;  Pred. No. 16;
  Matches   69;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy    183 cgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatcccctgcatcacc 242
          | | || | || |  ||  |     ||  || | || || || ||| ||||| || ||||
Db  14789 CCCCGCCAATATCGTTCCCAAACTGACACCTCCCGCGGCGGCGGCAGCCCCTCCAGCACC 14730

Qy    243 ccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggc 296
              ||| | || || ||   || |   || ||||| |||||||| | | | ||
Db  14729 GGAAGTTGCTGCACCGGCAGCGGTACTTGCTCCTCCAGCCGCGGCTCCACCCGC 14676


RESULT  15
US-09-739-449-217
; Sequence 217, Application US/09739449
; GENERAL INFORMATION:
;  APPLICANT: Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
;  FILE REFERENCE: 38-10(15490)C
;  CURRENT APPLICATION NUMBER: US/09/739,449
;  CURRENT FILING DATE:  2000-12-19
;  PRIOR APPLICATION NUMBER: US 09/514,000
;  PRIOR FILING DATE: 2000-02-23
;  NUMBER OF SEQ ID NOS: 13351
; SEQ ID NO 217
;   LENGTH: 397803
;   TYPE: DNA
;   ORGANISM: Agrobacterium tumefaciens
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(397803)
;   OTHER INFORMATION: unsure at all n locations
US-09-739-449-217


  Query Match             4.3%;  Score 42;  DB 29;  Length 397803;
  Best Local Similarity   60.5%;  Pred. No. 23;
  Matches   69;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy    183 cgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatcccctgcatcacc 242
          | | || | || |  ||  |     ||  || | || || || ||| ||||| || ||||
Db  66682 ccccgccaatatcgttcccaaactgacacctcccgcggcggcggcagccccctccagcacc 66741

Qy    243 ccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggc 296
              ||| | || || ||   || |   || ||||| |||||||| | | | ||
Db  66742 ggaagttgctgcaccggcagcggtacttgctcctccagccgcggctccacccgc 66795


Search completed: September 21, 2001, 02:19:42
Job time: 30454 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: September 21, 2001, 02:35:38 ; Search time 1035.34 Seconds
(without alignments)
1881.359 Million cell updates/sec

Title: US-09-138-735-1_COPY_1232_2207
Perfect score: 976
Sequence: 1 caggtacagcgtaacggctt.....tgtgaacgaggccattatgt 976

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1741792 seqs, 997871973 residues

Total number of hits satisfying chosen parameters: 3483584

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_New:*
1: /cgn1_7/ptodata/2/pna/PCT_NEW_COMB.seq:*
2: /cgn1_7/ptodata/2/pna/US06_NEW_COMB.seq:*
3: /cgn1_7/ptodata/2/pna/US07_NEW_COMB.seq:*
4: /cgn1_7/ptodata/2/pna/US08_NEW_COMB.seq:*
5: /cgn1_7/ptodata/2/pna/US09_NEW_COMB.seq:*
6: /cgn1_7/ptodata/2/pna/US09_NEW_COMB.seq1:*
7: /cgn1_7/ptodata/2/pna/US09_NEW_COMB.seq2:*
8: /cgn1_7/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	976	100.0	3402	6	US-09-138-735-1	Sequence 1, Appli
	2	42	4.3	397803	6	US-09-803-110-217	Sequence 217, App
c	3	40.4	4.1	477	7	US-09-684-843A-1	Sequence 1, Appli
	4	38.8	4.0	657	7	US-09-922-480-3	Sequence 3, Appli
	5	38.8	4.0	657	7	US-09-923-236-3	Sequence 3, Appli
	6	38.8	4.0	657	7	US-09-922-469-3	Sequence 3, Appli
	7	38.6	4.0	411	7	US-09-902-540-5862	Sequence 5862, Ap
	8	38.6	4.0	1209	7	US-09-902-540-147	Sequence 147, App
	9	38.2	3.9	495	8	US-60-253-651-24827	Sequence 24827, A
	10	38.2	3.9	670	8	US-60-253-651-24775	Sequence 24775, A
	11	37.8	3.9	405	8	US-60-253-654-1232	Sequence 1232, Ap
	12	37.8	3.9	405	8	US-60-255-592-1232	Sequence 1232, Ap
	13	37.8	3.9	578	8	US-60-253-654-18403	Sequence 18403, A
	14	37.8	3.9	578	8	US-60-255-592-18403	Sequence 18403, A
	15	37.8	3.9	608	8	US-60-253-654-18342	Sequence 18342, A
	16	37.8	3.9	608	8	US-60-255-592-18342	Sequence 18342, A
	17	37.4	3.8	1039	8	US-60-278-561-9647	Sequence 9647, Ap
	18	37.4	3.8	6712	8	US-60-278-232-2818	Sequence 2818, Ap
c	19	37.2	3.8	522	6	US-09-823-241-9418	Sequence 9418, Ap
c	20	37.2	3.8	2406	6	US-09-760-466-345	Sequence 345, App
	21	37.2	3.8	6483	7	US-09-880-107-3699	Sequence 3699, Ap
c	22	37.2	3.8	10733	7	US-09-902-540-1021	Sequence 1021, Ap
	23	36.8	3.8	435	5	US-09-842-827-1975	Sequence 1975, Ap
	24	36.8	3.8	435	7	US-09-912-292-40706	Sequence 40706, A
	25	36.8	3.8	621	7	US-09-840-424-56	Sequence 56, Appl
	26	36.6	3.8	30783	7	US-09-902-540-1258	Sequence 1258, Ap
	27	36.4	3.7	819	7	US-09-902-540-8196	Sequence 8196, Ap
c	28	36.4	3.7	1020	1	PCT-US01-08631-12729	Sequence 12729, A
	29	36.4	3.7	7190	7	US-09-902-540-843	Sequence 843, App
	30	36.2	3.7	389	6	US-09-652-125A-2957	Sequence 2957, Ap
c	31	36.2	3.7	692	6	US-09-866-555-15847	Sequence 15847, A
c	32	36.2	3.7	2405	1	PCT-US01-08631-9988	Sequence 9988, Ap
	33	36	3.7	334	7	US-09-850-147-15625	Sequence 15625, A
	34	36	3.7	483	6	US-09-795-300-42	Sequence 42, Appl
	35	36	3.7	488	7	US-09-834-366-26649	Sequence 26649, A
	36	36	3.7	505	5	US-09-846-039-59	Sequence 59, Appl
	37	36	3.7	592	8	US-60-253-654-24123	Sequence 24123, A
	38	36	3.7	592	8	US-60-255-592-24123	Sequence 24123, A
	39	36	3.7	1448	5	US-09-925-298-318	Sequence 318, App
	40	36	3.7	16080	7	US-09-724-568-48	Sequence 48, Appl
	41	36	3.7	16080	7	US-09-724-569-48	Sequence 48, Appl
	42	36	3.7	16080	7	US-09-724-571-48	Sequence 48, Appl
	43	35.8	3.7	446	7	US-09-912-292-19829	Sequence 19829, A
	44	35.6	3.6	299	7	US-09-859-490-2340	Sequence 2340, Ap
c	45	35.6	3.6	456	6	US-09-904-809-19195	Sequence 19195, A

ALIGNMENTS

RESULT 1
US-09-138-735-1
; Sequence 1, Application US/09138735
; GENERAL INFORMATION:
; APPLICANT: PARANHOS-BACCALA, Glaucia
; APPLICANT: LESENECHAL, Mylene
; APPLICANT: JOLIVET, Michel
; TITLE OF INVENTION: TRYPANOSOMA CRUZI ANTIGEN, GENE ENCODING THEREFOR AND METHODS
; TITLE OF INVENTION: DETECTING AND TREATING CHAGAS DISEASE
; FILE REFERENCE: WPB 36400B
; CURRENT APPLICATION NUMBER: US/09/138,735
; CURRENT FILING DATE: 1998-08-24
; PRIOR APPLICATION NUMBER: US 08/480,917
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: FR 94/10132
; PRIOR FILING DATE: 1994-08-12
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 3402
; TYPE: DNA
; ORGANISM: Trypanosoma cruzi
US-09-138-735-1

Query Match 100.0%; Score 976; DB 6; Length 3402;
Best Local Similarity 100.0%; Pred. No. 1.6e-228;
Matches 976; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 caggtacagcgtaacggcttttgcttcaatcgtacagccgacggtagctgcgtcctggct 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1232 caggtacagcgtaacggcttttgcttcaatcgtacagccgacggtagctgcgtcctggct 1291

Qy      61 gacatgtcgattcgattgacgatcttccatctccggtcctcccgcagggaagaacagcag 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1292 gacatgtcgattcgattgacgatcttccatctccggtcctcccgcagggaagaacagcag 1351

Qy     121 ccaggccaaaaaacatcggtagtggcgacggcgaaaccggggtgtgtgtcctcgggcact 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1352 ccaggccaaaaaacatcggtagtggcgacggcgaaaccggggtgtgtgtcctcgggcact 1411

Qy     181 gacgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatcccctgcatca 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1412 gacgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatcccctgcatca 1471

Qy     241 ccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggctgag 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1472 ccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggctgag 1531
```

```
Qy     301 ccgcacgtggggagcaagatcattgctaatctagtgaatcagctggggattaatgtcacc 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1532 ccgcacgtggggagcaagatcattgctaatctagtgaatcagctggggattaatgtcacc 1591

Qy     361 caaaggagcgtcgtcagcactggagcgccggccacgacgaggtctacggcggtgacgtcc 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1592 caaaggagcgtcgtcagcactggagcgccggccacgacgaggtctacggcggtgacgtcc 1651

Qy     421 acgactaccgccccgcagcgaacaagtccatacgggcacaatggccgacctgtgacggct 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1652 acgactaccgccccgcagcgaacaagtccatacgggcacaatggccgacctgtgacggct 1711

Qy     481 ggattggtggcagctaatagtggtgccagcgcggcctcgtctcccacagccgcggcgaaa 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1712 ggattggtggcagctaatagtggtgccagcgcggcctcgtctcccacagccgcggcgaaa 1771

Qy     541 ccaacaggagaagaaaaggcctccgcggcatgtgaaacgagctccgtggcgataaatgcg 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1772 ccaacaggagaagaaaaggcctccgcggcatgtgaaacgagctccgtggcgataaatgcg 1831

Qy     601 acacgcccggcgcttcacaacgcctctctcccgcaggcgccaacggatggcgttttggcg 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1832 acacgcccggcgcttcacaacgcctctctcccgcaggcgccaacggatggcgttttggcg 1891

Qy     661 gcagcagtataccagtcggagggcgaggttcatcagtcgctggagcggctggagtccgtc 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1892 gcagcagtataccagtcggagggcgaggttcatcagtcgctggagcggctggagtccgtc 1951

Qy     721 ataaccaacacgtctcgggttctgaagttgctccctgacaccattcgaagagaccatgaa 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1952 ataaccaacacgtctcgggttctgaagttgctccctgacaccattcgaagagaccatgaa 2011

Qy     781 caacttctgaatctgggtttagaggcacagatgacagagctgcagcagagccgtccaaca 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2012 caacttctgaatctgggtttagaggcacagatgacagagctgcagcagagccgtccaaca 2071

Qy     841 ccgcaaacacagccgagagacacaagctccgcgaaatcatccgtgtttgagacgtacacc 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2072 ccgcaaacacagccgagagacacaagctccgcgaaatcatccgtgtttgagacgtacacc 2131

Qy     901 cttgttctcattgcggattccctctctcgcaacatcacgaaggggtgaagcgtggtgtg 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2132 cttgttctcattgcggattccctctctcgcaacatcacgaaggggtgaagcgtggtgtg 2191

Qy     961 aacgaggccattatgt 976
           ||||||||||||||||
Db    2192 aacgaggccattatgt 2207
```

```
RESULT   2
US-09-803-110-217
; Sequence 217, Application US/09803110
; GENERAL INFORMATION:
;  APPLICANT: Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
;  FILE REFERENCE: 38-10(15490)D
;  CURRENT APPLICATION NUMBER: US/09/803,110
;  CURRENT FILING DATE:  2001-03-12
;  PRIOR APPLICATION NUMBER: US 09/739,449
;  PRIOR FILING DATE: 2000-12-19
;  PRIOR APPLICATION NUMBER: US 09/514,000
;  PRIOR FILING DATE: 2000-02-23
;  PRIOR APPLICATION NUMBER: US 60/168,139
;  PRIOR FILING DATE: 1999-12-01
;  NUMBER OF SEQ ID NOS: 13351
; SEQ ID NO 217
;   LENGTH: 397803
;   TYPE: DNA
;   ORGANISM: Agrobacterium tumefaciens
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(397803)
;   OTHER INFORMATION: unsure at all n locations
US-09-803-110-217


  Query Match             4.3%;  Score 42;  DB 6;  Length 397803;
  Best Local Similarity   60.5%;  Pred. No. 2.5;
  Matches   69;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy     183 cgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatcccctgcatcacc 242
           | | || | || |  ||  |     ||  || | || || || ||| ||||| || ||||
Db   66682 ccccgccaatatcgttcccaaactgacacctcccgcggcggcggcagcccctccagcacc 66741

Qy     243 ccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggc 296
               ||| | || || ||   || |   || ||||| |||||||| | | | ||
Db   66742 ggaagttgctgcaccggcagcggtacttgctcctccagccgcggctccacccgc 66795


RESULT   3
US-09-684-843A-1/c
; Sequence 1, Application US/09684843A
; GENERAL INFORMATION:
;  APPLICANT: Ranum et al.
;  TITLE OF INVENTION: SCA7 GENE AND METHODS OF USE
;  FILE REFERENCE: Regents of the University of Minnesota
;  CURRENT APPLICATION NUMBER: US/09/684,843A
;  CURRENT FILING DATE:  2000-10-06
;  PRIOR APPLICATION NUMBER: 60/056,170
;  PRIOR FILING DATE: 1997-08-19
;  PRIOR APPLICATION NUMBER: 09/135,994
;  PRIOR FILING DATE: 1998-08-18
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 477
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-684-843A-1


  Query Match             4.1%;  Score 40.4;  DB 7;  Length 477;
  Best Local Similarity   53.9%;  Pred. No. 1.7;
  Matches   83;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy     145 gcgacggcgaaaccggggtgtgtgtcctcgggcactgacgcggcgagtagcagtcatacc 204
           | | |||||    ||||   ||   || | ||  |||  | ||||        |  | |
Db     334 GTGGCGGCGGTGGCGGGTGCTGCTGCCGCTGGGGCTGCGGAGGCGGCGGCTGCTGCTGCT 275

Qy     205 aatacgacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaag 264
             | |  || |||| ||||||||  |  ||||  |  |  ||| | | ||  | ||
Db     274 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 215

Qy     265 gcagccgcgcctcctgccgcggcgcgatcggctg 298
           || || ||  || |||| || ||     | ||||
Db     214 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 181


RESULT   4
US-09-922-480-3
; Sequence 3, Application US/09922480
; GENERAL INFORMATION:
;  APPLICANT: Sheppard, Paul O.
;  APPLICANT:  Adler, David A.
;  TITLE OF INVENTION: SECRETED SALIVARY ZSIG63 POLYPEPTIDE
;  FILE REFERENCE: 97-71
;  CURRENT APPLICATION NUMBER: US/09/922,480
;  CURRENT FILING DATE:  2001-08-03
;  PRIOR APPLICATION NUMBER: US 60/124,820
;  PRIOR FILING DATE: 1999-03-17
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
;   LENGTH: 657
```

```
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Degenerate polynucleotide sequence for zsig63
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(657)
;   OTHER INFORMATION: n = A,T,C or G
US-09-922-480-3


  Query Match             4.0%;  Score 38.8;  DB 7;  Length 657;
  Best Local Similarity   31.4%;  Pred. No. 4.5;
  Matches   89;  Conservative   17;  Mismatches  177;  Indels    0;  Gaps    0;

Qy    157 ccggggtgtgtgtcctcgggcactgacgcggcgagtagcagtcataccaatacgacttct 216
          ||  :  |  | | ::: |   | |  ||  |          |  | :    |     |
Db    340 ccnwsnmgnttyttywsngcngcngcngcnccngcngcnccnccnathgcngcngarccn 399

Qy    217 gccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaaggcagccgcgcct 276
          || || || ||  | :   |  |  | || ||  | ||   : |    || |  || ||
Db    400 gcngcngcngcnccnytnacngcnacnccngtngcngcngarccngcngcnggngcnccn 459

Qy    277 cctgccgcggcgcgatcggctgagccgcacgtggggagcaagatcattgctaatctagtg 336
            || || | :|   | || ||: | |  || ||     |:      ||  |
Db    460 gtngcngcngarccngcngcngargcnccngtnggngcngarccngcngcngcngargcncn 519

Qy    337 aatcagctggggattaatgtcacccaaaggagcgtcgtcagcactggagcgccggccacg 396
              | :      |              |  ||     | |  ||   :|   |·
Db    520 gtngcngcngarccngcngcngargcnccngtnggngtngarccngcngcngcngargarccn 579

Qy    397 acgaggtctacggcggtgacgtccacgactaccgccccgcagc 439
          ::     |   : | |  ||  | | : |  | || ||  |:|
Db    580 wsnccngcngarccngcnacngcnaarccngcngcnccngarc 622


RESULT   5
US-09-923-236-3
; Sequence 3, Application US/09923236
; GENERAL INFORMATION:
;  APPLICANT: Sheppard, Paul O.
;  APPLICANT:  Adler, David A.
;  TITLE OF INVENTION: SECRETED SALIVARY ZSIG63 POLYPEPTIDE
;  FILE REFERENCE: 97-71
;  CURRENT APPLICATION NUMBER: US/09/923,236
;  CURRENT FILING DATE:  2001-08-03
;  PRIOR APPLICATION NUMBER: US 60/124,820
;  PRIOR FILING DATE: 1999-03-17
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
;   LENGTH: 657
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Degenerate polynucleotide sequence for zsig63
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(657)
;   OTHER INFORMATION: n = A,T,C or G
US-09-923-236-3


  Query Match             4.0%;  Score 38.8;  DB 7;  Length 657;
  Best Local Similarity   31.4%;  Pred. No. 4.5;
  Matches   89;  Conservative   17;  Mismatches  177;  Indels    0;  Gaps    0;

Qy    157 ccggggtgtgtgtcctcgggcactgacgcggcgagtagcagtcataccaatacgacttct 216
          ||  :  |  | | ::: |   | |  ||  |          |  | :    |     |
Db    340 ccnwsnmgnttyttywsngcngcngcngcnccngcngcnccnccnathgcngcngarccn 399

Qy    217 gccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaaggcagccgcgcct 276
          || || || ||  | :   |  |  | || ||  | ||   : |    || |  || ||
Db    400 gcngcngcngcnccnytnacngcnacnccngtngcngcngarccngcngcnggngcnccn 459

Qy    277 cctgccgcggcgcgatcggctgagccgcacgtggggagcaagatcattgctaatctagtg 336
            || || | :|   | || ||: | |  || ||     |:      ||  |
Db    460 gtngcngcngarccngcngcngargcnccngtnggngcngarccngcngcngcngargcncn 519

Qy    337 aatcagctggggattaatgtcacccaaaggagcgtcgtcagcactggagcgccggccacg 396
              | :      |              |  ||     | |  ||   :|   |
Db    520 gtngcngcngarccngcngcngargcnccngtnggngtngarccngcngcngargarccn 579

Qy    397 acgaggtctacggcggtgacgtccacgactaccgccccgcagc 439
          ::     |   : | |  ||  | | : |  | || ||  |:|
Db    580 wsnccngcngarccngcnacngcnaarccngcngcnccngarc 622


RESULT   6
US-09-922-469-3
; Sequence 3, Application US/09922469
; GENERAL INFORMATION:
;  APPLICANT: Sheppard, Paul O.
;  APPLICANT:  Adler, David A.
;  TITLE OF INVENTION: SECRETED SALIVARY ZSIG63 POLYPEPTIDE
;  FILE REFERENCE: 97-71
;  CURRENT APPLICATION NUMBER: US/09/922,469
;  CURRENT FILING DATE:  2001-08-03
;  PRIOR APPLICATION NUMBER: US 60/124,820
;  PRIOR FILING DATE: 1999-03-17
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
;   LENGTH: 657
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Degenerate polynucleotide sequence for zsig63
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(657)
;   OTHER INFORMATION: n = A,T,C or G
US-09-922-469-3


  Query Match             4.0%;  Score 38.8;  DB 7;  Length 657;
  Best Local Similarity   31.4%;  Pred. No. 4.5;
  Matches   89;  Conservative   17;  Mismatches  177;  Indels    0;  Gaps    0;

Qy    157 ccggggtgtgtgtcctcgggcactgacgcggcgagtagcagtcataccaatacgacttct 216
          ||  :  |  | | ::: |   | |  ||  |          |  | :    |     |
Db    340 ccnwsnmgnttyttywsngcngcngcngcnccngcngcnccnccnathgcngcngarccn 399

Qy    217 gccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaaggcagccgcgcct 276
          || || || ||  | :   |  |  | || ||  | ||   : |    || |  || ||
Db    400 gcngcngcngcnccnytnacngcnacnccngtngcngcngarccngcngcnggngcnccn 459

Qy    277 cctgccgcggcgcgatcggctgagccgcacgtggggagcaagatcattgctaatctagtg 336
            || || | :|   | || ||: | |  || ||     |:      ||  |
Db    460 gtngcngcngarccngcngcngargcnccngtnggngcngarccngcngcngcngargcncn 519

Qy    337 aatcagctggggattaatgtcacccaaaggagcgtcgtcagcactggagcgccggccacg 396
              | :      |              |  ||     | |  ||   :|   |
Db    520 gtngcngcngarccngcngcngargcnccngtnggngtngarccngcngcngargarccn 579

Qy    397 acgaggtctacggcggtgacgtccacgactaccgccccgcagc 439
          ::     |   : | |  ||  | | : |  | || ||  |:|
Db    580 wsnccngcngarccngcnacngcnaarccngcngcnccngarc 622


RESULT   7
US-09-902-540-5862
; Sequence 5862, Application US/09902540
; GENERAL INFORMATION:
;  APPLICANT: Goldman, Barry S.
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Wiegand, Roger C.
;  TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
```

```
;  FILE REFERENCE: 38-10(15849)B
;  CURRENT APPLICATION NUMBER: US/09/902,540
;  CURRENT FILING DATE:  2001-07-10
;  PRIOR APPLICATION NUMBER: 60/217,883
;  PRIOR FILING DATE: 2000-07-10
;  NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 5862
;   LENGTH: 411
;   TYPE: DNA
;   ORGANISM: Myxococcus xanthus
US-09-902-540-5862


  Query Match             4.0%;  Score 38.6;  DB 7;  Length 411;
  Best Local Similarity   52.7%;  Pred. No. 4.7;
  Matches  107;  Conservative    0;  Mismatches   94;  Indels    2;  Gaps    1;

Qy    215 ctgccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaaggcagccgcgc 274
          ||||  | |||||| | | | | |||||  | | |   ||| ||    | || |  | ||
Db     58 ctgctccggctgcaccgcattcgtcacctgccggtggtgcggcaagagaagctggtgggc 117

Qy    275 ctcctgccgcggcgcgatcggctgagccgcacgtggggagcaagatcattgctaatctag 334
          ||| |  | |  ||||| | |||| ||    ||  |  | | |   ||  |         |
Db    118 ctcatcacccaccgcgacctgctgcgc--gccgccgcaacccacgccacggacccggccg 175

Qy    335 tgaatcagctggggattaatgtcacccaaaggagcgtcgtcagcactggagcgccggcca 394
          | | | |||  ||      | || |   | | ||| |||    || |     |||| ||
Db    176 cgcagccgctctgggccgcggacatcatgacgcgcgacgtgcagacggtgcgcccggaca 235

Qy    395 cgacgaggtctacggcggtgacg 417
          || ||  |     ||||||||||
Db    236 cgccgctgcgccgggcggtgacg 258


RESULT   8
US-09-902-540-147
; Sequence 147, Application US/09902540
; GENERAL INFORMATION:
;  APPLICANT: Goldman, Barry S.
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Wiegand, Roger C.
;  TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
;  FILE REFERENCE: 38-10(15849)B
;  CURRENT APPLICATION NUMBER: US/09/902,540
;  CURRENT FILING DATE:  2001-07-10
;  PRIOR APPLICATION NUMBER: 60/217,883
;  PRIOR FILING DATE: 2000-07-10
;  NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 147
;   LENGTH: 1209
;   TYPE: DNA
;   ORGANISM: Myxococcus xanthus
US-09-902-540-147


  Query Match             4.0%;  Score 38.6;  DB 7;  Length 1209;
  Best Local Similarity   52.7%;  Pred. No. 5.7;
  Matches  107;  Conservative    0;  Mismatches   94;  Indels    2;  Gaps    1;

Qy    215 ctgccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaaggcagccgcgc 274
          ||||  | |||||| | | | | |||||  | | |   ||| ||    | || |  | ||
Db    285 ctgctccggctgcaccgcattcgtcacctgccggtggtgcggcaagagaagctggtgggc 344

Qy    275 ctcctgccgcggcgcgatcggctgagccgcacgtggggagcaagatcattgctaatctag 334
          ||| |  | |  ||||| | |||| ||    ||  |  | | |   ||  |         |
Db    345 ctcatcacccaccgcgacctgctgcgc--gccgccgcaacccacgccacggacccggccg 402

Qy    335 tgaatcagctggggattaatgtcacccaaaggagcgtcgtcagcactggagcgccggcca 394
          | | | |||  ||      | || |   | | ||| |||    || |     |||| ||
Db    403 cgcagccgctctgggccgcggacatcatgacgcgcgacgtgcagacggtgcgcccggaca 462

Qy    395 cgacgaggtctacggcggtgacg 417
          || ||  |     ||||||||||
Db    463 cgccgctgcgccgggcggtgacg 485


RESULT   9
US-60-253-651-24827
; Sequence 24827, Application US/60253651
; GENERAL INFORMATION:
;  APPLICANT: Glenn, Mattew
;  APPLICANT:  Grigor, Murray R
;  TITLE OF INVENTION: Compositions isolated from bovine tissue
;  TITLE OF INVENTION:  and methods for their use.
;  FILE REFERENCE: 1055P3
;  CURRENT APPLICATION NUMBER: US/60/253,651
;  CURRENT FILING DATE:  2000-11-28
;  NUMBER OF SEQ ID NOS: 27858
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 24827
;   LENGTH: 495
;   TYPE: DNA
;   ORGANISM: Bovine
US-60-253-651-24827


  Query Match             3.9%;  Score 38.2;  DB 8;  Length 495;
  Best Local Similarity   49.7%;  Pred. No. 6;
  Matches   97;  Conservative    0;  Mismatches   98;  Indels    0;  Gaps    0;

Qy    106 agggaagaacagcagccaggccaaaaaaacatcggtagtggcgacggcgaaaccggggtgt 165
          |||||||  | |  | | |  |  |    |  |||  | ||| | |||| ||     |
Db    128 agggaagttccggcggcggctcctacggcggcgggcgcggctatggcggaagttccggcg 187

Qy    166 gtgtcctcgggcactgacgcggcgagtagcagtcataccaatacgacttctgccgctgct 225
          | | |  |||   ||   ||||||              |   |||| |  | |  ||| |
Db    188 gcggcggcggcggctacagcggcggaagctccagcggcggctacggcggcggaagctcca 247

Qy    226 gcatcccctgcatcacccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcg 285
          ||     | ||  |  |    ||| | ||| |||| | ||| || |   |||| |  |||
Db    248 gcggtggcggccacggcggtagttccggcggcagctacggcggcggaagctccagtggcg 307

Qy    286 gcgcgatcggctgag 300
          | |    |||| | |
Db    308 gtggccacggcggcg 322


RESULT  10
US-60-253-651-24775
; Sequence 24775, Application US/60253651
; GENERAL INFORMATION:
;  APPLICANT: Glenn, Mattew
;  APPLICANT:  Grigor, Murray R
;  TITLE OF INVENTION: Compositions isolated from bovine tissue
;  TITLE OF INVENTION:  and methods for their use.
;  FILE REFERENCE: 1055P3
;  CURRENT APPLICATION NUMBER: US/60/253,651
;  CURRENT FILING DATE:  2000-11-28
;  NUMBER OF SEQ ID NOS: 27858
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 24775
;   LENGTH: 670
;   TYPE: DNA
;   ORGANISM: Bovine
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(670)
;   OTHER INFORMATION: n = A,T,C or G
US-60-253-651-24775


  Query Match             3.9%;  Score 38.2;  DB 8;  Length 670;
  Best Local Similarity   49.7%;  Pred. No. 6.4;
  Matches   97;  Conservative    0;  Mismatches   98;  Indels    0;  Gaps    0;
```

```
Qy    106 agggaagaacagcagccaggccaaaaaacatcggtagtggcgacggcgaaaccggggtgt 165
          ||||||| | |  |.| |  |  |   |  |||  | ||| | |||| ||     |
Db    332 agggaagttccggcggcggctcctacggcggcgggcgcggctatggcggaagttccggcg 391

Qy    166 gtgtcctcgggcactgacgcggcgagtagcagtcataccaatacgacttctgccgctgct 225
          | | |  |||   ||   ||||||              |   |||| |  | |  ||| |
Db    392 gcggcggcggcggctacggcggcggaagctccagcggcggctacggcggcggaagctcca 451

Qy    226 gcatcccctgcatcacccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcg 285
          ||     | ||  |  |    ||| | ||| |||| | ||| || |   |||| |  |||
Db    452 gcggtggcggccacggcggtagttccggcggcagctacggcggcggaagctccagtggcg 511

Qy    286 gcgcgatcggctgag 300
          | |    |||| | |
Db    512 gtggccacggcggcg 526


RESULT  11
US-60-253-654-1232
; Sequence 1232, Application US/60253654
; GENERAL INFORMATION:
;  APPLICANT: Glenn, Matthew
;  APPLICANT:  Norriss, Michael G.
;  TITLE OF INVENTION: Compositions isolated from forage
;  TITLE OF INVENTION:  grasses and methods for their use.
;  FILE REFERENCE: 1056P
;  CURRENT APPLICATION NUMBER: US/60/253,654
;  CURRENT FILING DATE:  2000-11-28
;  NUMBER OF SEQ ID NOS: 32581
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1232
;   LENGTH: 405
;   TYPE: DNA
;   ORGANISM: Lolium perenne
US-60-253-654-1232


  Query Match             3.9%;  Score 37.8;  DB 8;  Length 405;
  Best Local Similarity   46.2%;  Pred. No. 7.3;
  Matches  126;  Conservative    0;  Mismatches  147;  Indels    0;  Gaps    0;

Qy    153 gaaaccggggtgtgtgtcctcgggcactgacgcggcgagtagcagtcataccaatacgac 212
          | |  ||| |    || || ||   || | |||| |||  |  |   |  |||   |
Db     83 ggatacggagcacctgcccccgcctacggccgcgacgacgactacggacgccacgcccca 142

Qy    213 ttctgccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaaggcagccgc 272
            |  |  || | ||  ||||  || | ||  |  ||  |  || |  |  ||| |
Db    143 gcccccgcctacggcgcccccgccaccgcctacggtcgcgacgacgactacggacgccac 202

Qy    273 gcctcctgccgcggcgcgatcggctgagccgcacgtggggagcaagatcattgctaatct 332
          ||| ||  ||  | |     |||| |   ||     ||  |       ||       ||
Db    203 gcccccgcccccggcctacggcggcggtcgcgacgacggctacggtcgccacgagccggct 262

Qy    333 agtgaatcagctggggattaatgtcacccaaaggagcgtcgtcagcactggagcgccggc 392
            |  |  |  || |      |  ||  |       | |        |   ||| ||||
Db    263 ccggcctacgggggaggcggcggctacggagccccttgcccagcagcttacggcggcggc 322

Qy    393 cacgacgaggtctacggcggtgacgtccacgac 425
          | |||||| | |||||||  |  ||| ||  | |
Db    323 cgcgacgacggctacggccgccacgccccggcc 355


RESULT  12
US-60-255-592-1232
; Sequence 1232, Application US/60255592
; GENERAL INFORMATION:
;  APPLICANT: Glenn, Matthew
;  APPLICANT:  Norriss, Michael G.
;  TITLE OF INVENTION: Compositions isolated from forage
;  TITLE OF INVENTION:  grasses and methods for their use.
;  FILE REFERENCE: 1056P2
;  CURRENT APPLICATION NUMBER: US/60/255,592
;  CURRENT FILING DATE:  2000-12-14
;  NUMBER OF SEQ ID NOS: 32581
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1232
;   LENGTH: 405
;   TYPE: DNA
;   ORGANISM: Lolium perenne
US-60-255-592-1232


  Query Match             3.9%;  Score 37.8;  DB 8;  Length 405;
  Best Local Similarity   46.2%;  Pred. No. 7.3;
  Matches  126;  Conservative    0;  Mismatches  147;  Indels    0;  Gaps    0;

Qy    153 gaaaccggggtgtgtgtcctcgggcactgacgcggcgagtagcagtcataccaatacgac 212
          | |  ||| |    || || ||   || | |||| |||  |  |   |  |||   |
Db     83 ggatacggagcacctgcccccgcctacggccgcgacgacgactacggacgccacgcccca 142

Qy    213 ttctgccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaaggcagccgc 272
            |  |  || | ||  ||||  || | ||  |  ||  |  || |  |  ||| |
Db    143 gcccccgcctacggcgcccccgccaccgcctacggtcgcgacgacgactacggacgccac 202

Qy    273 gcctcctgccgcggcgcgatcggctgagccgcacgtggggagcaagatcattgctaatct 332
          ||| ||  ||  | |     |||| |   ||     ||  |       ||       ||
Db    203 gcccccgcccccggcctacggcggcggtcgcgacgacggctacggtcgccacgagccggct 262

Qy    333 agtgaatcagctggggattaatgtcacccaaaggagcgtcgtcagcactggagcgccggc 392
            |  |  |  || |      |  ||  |       | |        |   ||| ||||
Db    263 ccggcctacgggggaggcggcggctacggagccccttgcccagcagcttacggcggcggc 322

Qy    393 cacgacgaggtctacggcggtgacgtccacgac 425
          | |||||| | |||||||  |  ||| ||  | |
Db    323 cgcgacgacggctacggccgccacgccccggcc 355


RESULT  13
US-60-253-654-18403
; Sequence 18403, Application US/60253654
; GENERAL INFORMATION:
;  APPLICANT: Glenn, Matthew
;  APPLICANT:  Norriss, Michael G.
;  TITLE OF INVENTION: Compositions isolated from forage
;  TITLE OF INVENTION:  grasses and methods for their use.
;  FILE REFERENCE: 1056P
;  CURRENT APPLICATION NUMBER: US/60/253,654
;  CURRENT FILING DATE:  2000-11-28
;  NUMBER OF SEQ ID NOS: 32581
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18403
;   LENGTH: 578
;   TYPE: DNA
;   ORGANISM: Festuca arundinaceae
US-60-253-654-18403


  Query Match             3.9%;  Score 37.8;  DB 8;  Length 578;
  Best Local Similarity   64.0%;  Pred. No. 7.8;
  Matches   57;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy    209 cgacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaaggcag 268
          ||||  |   |||| ||  | || ||  || | || || | |  |||| |||||   ||||
Db    330 cgacgtgctccgccgcctcgtcgccaacaccgccaccggctcgagcggcagccccgcag 389

Qy    269 ccgcgcctcctgccgcggcgcgatcggct 297
          ||  ||  |  | | || |||| |  |||||
Db    390 ccacgacggcggacggggcgagggcggct 418


RESULT  14
US-60-255-592-18403
; Sequence 18403, Application US/60255592
; GENERAL INFORMATION:
;  APPLICANT: Glenn, Matthew
;  APPLICANT:  Norriss, Michael G.
```

```
;  TITLE OF INVENTION: Compositions isolated from forage
;  TITLE OF INVENTION:  grasses and methods for their use.
;  FILE REFERENCE: 1056P2
;  CURRENT APPLICATION NUMBER: US/60/255,592
;  CURRENT FILING DATE:  2000-12-14
;  NUMBER OF SEQ ID NOS: 32581
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18403
;   LENGTH: 578
;   TYPE: DNA
;   ORGANISM: Festuca arundinaceae
US-60-255-592-18403


  Query Match             3.9%;  Score 37.8;  DB 8;  Length 578;
  Best Local Similarity   64.0%;  Pred. No. 7.8;
  Matches   57;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy     209 cgacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaaggcag 268
           |||| |   |||| ||  | || ||  || | || || | |  |||| |||||   ||||
Db     330 cgacgtgctccgccgcctcgtcgccaacaccgccaccggctcgagcggcagccccccgcag 389

Qy     269 ccgcgcctcctgccgcggcgcgatcggct 297
           || || |  | | || |||| |  |||||
Db     390 ccacgacggcggacggggcgagggcggct 418


RESULT  15
US-60-253-654-18342
; Sequence 18342, Application US/60253654
; GENERAL INFORMATION:
;  APPLICANT: Glenn, Matthew
;  APPLICANT:  Norriss, Michael G.
;  TITLE OF INVENTION: Compositions isolated from forage
;  TITLE OF INVENTION:  grasses and methods for their use.
;  FILE REFERENCE: 1056P
;  CURRENT APPLICATION NUMBER: US/60/253,654
;  CURRENT FILING DATE:  2000-11-28
;  NUMBER OF SEQ ID NOS: 32581
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18342
;   LENGTH: 608
;   TYPE: DNA
;   ORGANISM: Festuca arundinaceae
US-60-253-654-18342


  Query Match             3.9%;  Score 37.8;  DB 8;  Length 608;
  Best Local Similarity   64.0%;  Pred. No. 7.9;
  Matches   57;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy     209 cgacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaaggcag 268
           |||| |   |||| ||  | || ||  || | || || | |  |||| |||||   ||||
Db     330 cgacgtgctccgccgcctcgtcgccaacaccgccaccggctcgagcggcagccccccgcag 389

Qy     269 ccgcgcctcctgccgcggcgcgatcggct 297
           || || |  | | || |||| |  |||||
Db     390 ccacgacggcggacggggcgagggcggct 418


Search completed: September 21, 2001, 02:39:09
Job time: 15035 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: September 21, 2001, 00:34:56 ; Search time 3479.37 Seconds
(without alignments)
80.020 Million cell updates/sec

Title: US-09-138-735-8
Perfect score: 18
Sequence: 1 tgcagcagcggcagaagt 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*
44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vi1:*
59: gb_vi2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	18	100.0	18	9	A48917	A48917 Sequence 8
	2	18	100.0	18	9	AR047926	AR047926 Sequence
	3	18	100.0	1589	10	AX086831	AX086831 Sequence
	4	18	100.0	1589	93	HSM801826	AL136858 Homo sapi
c	5	18	100.0	3402	9	A48910	A48910 Sequence 1
c	6	18	100.0	3402	9	AR047920	AR047920 Sequence
c	7	18	100.0	3402	96	TCU24190	U24190 Trypanosoma
	8	18	100.0	11873	9	AR083116	AR083116 Sequence

	9	18	100.0	11878	9	AR083115	AR083115 Sequence
	10	18	100.0	11883	9	AR083114	AR083114 Sequence
	11	17	94.4	915	77	AC085350	AC085350 Giardia i
c	12	17	94.4	920	71	AC034857	AC034857 Giardia i
	13	17	94.4	1009	71	AC029282	AC029282 Giardia i
	14	17	94.4	1671	53	AF309358	AF309358 Hordeum v
c	15	17	94.4	4545	94	AB030448S1	AB030448 Mus muscu
c	16	17	94.4	5851	94	AF152533S1	AF152533 Mus muscu
c	17	17	94.4	6074	94	MUSP3VPR1	AF098866 Mus sp. 1
c	18	17	94.4	65174	69	AC025214	AC025214 Homo sapi
	19	17	94.4	91470	12	AC011438	AC011438 Genomic s
	20	17	94.4	110000	84	LMFLCHR36_09	Continuation (10 o
	21	17	94.4	157981	85	AC005184	AC005184 Homo sapi
	22	17	94.4	167590	80	AL356221	AL356221 Homo sapi
	23	17	94.4	171539	81	AL512788	AL512788 Homo sapi
	24	17	94.4	197871	73	AC068053	AC068053 Homo sapi
	25	17	94.4	200033	93	HSA251973	AJ251973 Homo sapi
c	26	16.4	91.1	160	54	HSPA33G6	Z94264 H.sapiens f
	27	16.4	91.1	652	2	BACNPRR1	M27555 Bacillis su
	28	16.4	91.1	876	9	A14600	A14600 Nucleotide
	29	16.4	91.1	876	10	E01244	E01244 DNA sequenc
c	30	16.4	91.1	1399	94	AF192493	AF192493 Rattus no
c	31	16.4	91.1	1406	94	AF192494	AF192494 Rattus no
c	32	16.4	91.1	2573	94	AF190798	AF190798 Rattus no
c	33	16.4	91.1	3259	91	BC001222	BC001222 Homo sapi
c	34	16.4	91.1	3436	89	AK001621	AK001621 Homo sapi
	35	16.4	91.1	5813	85	AB040944	AB040944 Homo sapi
	36	16.4	91.1	8014	63	AC014253	AC014253 Drosophil
c	37	16.4	91.1	14210	2	BSZ97025	Z97025 Bacillus su
c	38	16.4	91.1	16427	65	AC020116	AC020116 Drosophil
	39	16.4	91.1	36519	9	AR101859	AR101859 Sequence
c	40	16.4	91.1	36929	78	AC091077	AC091077 Homo sapi
	41	16.4	91.1	48443	4	AC005656	AC005656 Drosophil
	42	16.4	91.1	76452	65	AC020339	AC020339 Drosophil
	43	16.4	91.1	78260	63	AC015439	AC015439 Drosophil
c	44	16.4	91.1	104467	61	AC010067	AC010067 Drosophil
c	45	16.4	91.1	106377	85	AC004132	AC004132 Homo sapi

ALIGNMENTS

```
RESULT   1
A48917
LOCUS       A48917          18 bp    DNA             PAT       07-MAR-1997
DEFINITION  Sequence 8 from Patent WO9605312.
ACCESSION   A48917
VERSION     A48917.1  GI:2302576
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 18)
  AUTHORS   Paranhos-Baccala,G., Lesenechal,M. and Jolivet,M.
  TITLE     NOVEL TRYPANOSOMA CRUZI ANTIGEN, AND GENE CODING THEREFOR; THEIR
            APPLICATION TO THE DETECTION OF CHAGAS' DISEASE
  JOURNAL   Patent: WO 9605312-A 8 22-FEB-1996;
            BIO MERIEUX (FR)
COMMENT     Other publication CA 2173957 960222
            Other publication AU 3169195 960307
            Other publication FR 2723589 960216.
FEATURES             Location/Qualifiers
     source          1. .18
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT        5 a      4 c      7 g      2 t
ORIGIN


  Query Match           100.0%;  Score 18;  DB 9;  Length 18;
  Best Local Similarity  100.0%;  Pred. No. 5.8e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||||||||||||||||
Db      1 TGCAGCAGCGGCAGAAGT 18


RESULT   2
AR047926
LOCUS       AR047926        18 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 8 from patent US 5820864.
ACCESSION   AR047926
VERSION     AR047926.1  GI:5970269
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 18)
  AUTHORS   Paranhos-Baccala,G., Lesenechal,M. and Jolivet,M.
  TITLE     Trypanosoma cruzi antigen, gene encoding therefor and methods of
            detecting and treating chagas disease
  JOURNAL   Patent: US 5820864-A 8 13-OCT-1998;
FEATURES             Location/Qualifiers
     source          1. .18
                     /organism="unknown"
BASE COUNT        5 a      4 c      7 g      2 t
ORIGIN


  Query Match           100.0%;  Score 18;  DB 9;  Length 18;
  Best Local Similarity  100.0%;  Pred. No. 5.8e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||||||||||||||||
Db      1 TGCAGCAGCGGCAGAAGT 18


RESULT   3
AX086831
LOCUS       AX086831     1589 bp    DNA             PAT       09-MAR-2001
DEFINITION  Sequence 783 from Patent WO0112659.
ACCESSION   AX086831
VERSION     AX086831.1  GI:13276044
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1589)
  AUTHORS   Wiemann,S.
  TITLE     Human dna sequences
  JOURNAL   Patent: WO 0112659-A 783 22-FEB-2001;
            German Human Genome Project (DE)
FEATURES             Location/Qualifiers
     source          1. .1589
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      442 a    374 c    353 g    420 t
ORIGIN


  Query Match           100.0%;  Score 18;  DB 10;  Length 1589;
  Best Local Similarity  100.0%;  Pred. No. 2.6e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||||||||||||||||
Db    316 TGCAGCAGCGGCAGAAGT 333


RESULT   4
HSM801826
LOCUS       HSM801826    1589 bp    mRNA            PRI       10-MAR-2001
DEFINITION  Homo sapiens mRNA; cDNA DKFZp434N2435 (from clone DKFZp434N2435);
```

```
            complete cds.
ACCESSION   AL136858
VERSION     AL136858.1  GI:12053220
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1589)
  AUTHORS   Wiemann,S., Weil,B., Wellenreuther,R., Gassenhuber,J., Glassl,S.,
            Ansorge,W., Boecher,M., Bloecker,H., Bauersachs,S., Blum,H.,
            Lauber,J., Duesterhoeft,A., Beyer,A., Koehrer,K., Strack,N.,
            Mewes,H.W., Ottenwaelder,B., Obermaier,B., Tampe,J., Heubner,D.,
            Wambutt,R., Korn,B., Klein,M. and Poustka,A.
  TITLE     Toward a Catalog of Human Genes and Proteins: Sequencing and
            Analysis of 500 Novel Complete Protein Coding Human cDNAs
  JOURNAL   Genome Res. 11 (3), 422-435 (2001)
   PUBMED   11230166
REFERENCE   2  (bases 1 to 1589)
  AUTHORS   Poustka,A., Klein,M., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JAN-2000) MIPS, Am Klopferspitz 18a, D-82152
            Martinsried, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by DKFZ (German Cancer Research Center,
            Heidelberg/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            This clone is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
            information about the clone and the sequencing project is available
            at http://www.mips.biochem.mpg.de/proj/cDNA/.
FEATURES             Location/Qualifiers
     source          1. .1589
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DKFZp434N2435"
                     /tissue_type="testis"
                     /clone_lib="434 (synonym: htes3). Vector pSport1; host
                     DH10B; sites NotI + SalI"
                     /dev_stage="adult"
     gene            78. .1175
                     /gene="DKFZp434N2435"
     CDS             78. .1175
                     /gene="DKFZp434N2435"
                     /note="unknown protein"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAB66792.1"
                     /db_xref="GI:12053221"
                     /translation="MNVIYPLAVPKGRRLCCEVCEAPAERVCAACTVTYYCGVVHQKA
                     DWDSIHEKICQLLIPLRTSMPFYNSEEERQHGLQQLQQRQKYLIEFCYTIAQKYLFEG
                     KHEDAVPAALQSLRFRVKLYGLSSVELVPAYPLLAEASLGLGRIVQAEEYLFQAQWTV
                     LKSTDCSNATHSLLHRNLGLLYIAKKNYEEARYHLANDIYFASCAFGTEDIRTSGGYF
                     HLANIFYDLKKLDLADTLYTKVSEIWHAYLNNHYQVLSQAHIQQMDLLGKLFENDTGL
                     DEAQEAEAIRILTSILNIRESTSDKAPQKTIFVLKILVMLYYLMMNSSKAQEYGMRAL
                     SLAKEQQLDVHEQSTIQELLSLISTEDHPIT"
     polyA_signal    1561. .1566
     polyA_site      1579
BASE COUNT      442 a    374 c    353 g    420 t
ORIGIN
```

```
  Query Match             100.0%;  Score 18;  DB 93;  Length 1589;
  Best Local Similarity   100.0%;  Pred. No. 2.6e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||||||||||||||||
Db    316 TGCAGCAGCGGCAGAAGT 333
```

```
RESULT   5
A48910/c
LOCUS       A48910       3402 bp    DNA             PAT       07-MAR-1997
DEFINITION  Sequence 1 from Patent WO9605312.
ACCESSION   A48910
VERSION     A48910.1  GI:2302570
KEYWORDS    .
SOURCE      Trypanosoma cruzi.
  ORGANISM  Trypanosoma cruzi
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma; Schizotrypanum.
REFERENCE   1  (bases 1 to 3402)
  AUTHORS   Paranhos-Baccala,G., Lesenechal,M. and Jolivet,M.
  TITLE     NOVEL TRYPANOSOMA CRUZI ANTIGEN, AND GENE CODING THEREFOR; THEIR
            APPLICATION TO THE DETECTION OF CHAGAS' DISEASE
  JOURNAL   Patent: WO 9605312-A 1 22-FEB-1996;
            BIO MERIEUX (FR)
COMMENT     Other publication CA 2173957 960222
            Other publication AU 3169195 960307
            Other publication FR 2723589 960216.
FEATURES             Location/Qualifiers
     source          1. .3402
                     /organism="Trypanosoma cruzi"
                     /strain="G"
                     /db_xref="taxon:5693"
                     /dev_stage="EPIMASTIGOTE"
BASE COUNT      889 a    818 c    958 g    737 t
ORIGIN
```

```
  Query Match             100.0%;  Score 18;  DB 9;  Length 3402;
  Best Local Similarity   100.0%;  Pred. No. 2.3e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||||||||||||||||
Db   1459 TGCAGCAGCGGCAGAAGT 1442
```

```
RESULT   6
AR047920/c
LOCUS       AR047920     3402 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 1 from patent US 5820864.
ACCESSION   AR047920
VERSION     AR047920.1  GI:5970263
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 3402)
  AUTHORS   Paranhos-Baccala,G., Lesenechal,M. and Jolivet,M.
  TITLE     Trypanosoma cruzi antigen, gene encoding therefor and methods of
            detecting and treating chagas disease
  JOURNAL   Patent: US 5820864-A 1 13-OCT-1998;
FEATURES             Location/Qualifiers
     source          1. .3402
                     /organism="unknown"
BASE COUNT      888 a    821 c    956 g    737 t
ORIGIN
```

```
  Query Match             100.0%;  Score 18;  DB 9;  Length 3402;
  Best Local Similarity   100.0%;  Pred. No. 2.3e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||||||||||||||||
Db   1459 TGCAGCAGCGGCAGAAGT 1442
```

```
RESULT   7
TCU24190/c
LOCUS       TCU24190     3402 bp    mRNA            INV       04-AUG-1997
```

```
DEFINITION  Trypanosoma cruzi Tc40 antigen (Tc40) mRNA, complete cds.
ACCESSION   U24190
VERSION     U24190.1  GI:790645
KEYWORDS    .
SOURCE      Trypanosoma cruzi.
  ORGANISM  Trypanosoma cruzi
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma; Schizotrypanum.
REFERENCE   1  (bases 1 to 3402)
  AUTHORS   Lesenechal,M., Duret,L., Cano,M.I., Mortara,R.A., Jolivet,M.,
            Camargo,M.E., da Silveira,J.F. and Paranhos-Baccala,G.
  TITLE     Cloning and characterization of a gene encoding a novel
            immunodominant antigen of Trypanosoma cruzi
  JOURNAL   Mol. Biochem. Parasitol. 87 (2), 193-204 (1997)
  MEDLINE   97391123
REFERENCE   2  (bases 1 to 3402)
  AUTHORS   Lesenechal,M., Franco Da Silveira,J., Mortara,R.A., Duret,L.,
            Camargo,M.E., Jolivet,M. and Paranhos-Baccala,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-APR-1995) Mylene Lesenechal, BioMerieux, 22 rue Saint
            Jean de Dieu, Lyon 69007, France
FEATURES             Location/Qualifiers
     source          1. .3402
                     /organism="Trypanosoma cruzi"
                     /strain="G"
                     /db_xref="taxon:5693"
                     /dev_stage="epimastigote"
     misc_RNA        1. .35
                     /gene="Tc40"
                     /note="spliced leader"
     gene            1. .3402
                     /gene="Tc40"
     CDS             266. .3013
                     /gene="Tc40"
                     /codon_start=1
                     /product="Tc40 antigen"
                     /protein_id="AAC47657.1"
                     /db_xref="GI:790646"
                     /translation="MTVTVDLFNHAKPSNNEGRVWSVDAATFNEVPEAQRVLADSQFY
                     LAYTMKRRHVLRVVKRSNLLKGTVRAHSKPIHAVKFVNYRSNVAASAGKGEFFVWVVT
                     DETEASNGKPDLAARLTVKVYFKLQDPVTIPCFSFFINAESQRPDLLVLYETQAAILD
                     SSSLIERFDVESLEATLQRNCTTLRTLTQPVSENSLCSVGSGGWFTFTTEPTMVAACT
                     LRNRSTPSWACCEGEPVKALHLLDATVEENVSVLVAASTKGVYQWLLTGVAEPNLLRK
                     FVIDGSIVAMESSRETFAVFDDRKQLALVNMHSPHNFTCTHYMMPCQVQRNGFCFNRT
                     ADGSCVLADMSNRLTIFHLRSSRREEQQPGQKTSVVATAKPGCVSSGTDAASSSHTNT
                     TSAAAASPASPPVSAPAKAAAPPAAARSAEPHVGSKIIANLVNQLGINVTQRSVVSTG
                     APATTRSTAVTSTTTAPQRTSPYGHNGRPVTAGLVAANSGASAASSPTAAAKPTGEEK
                     ASAACETSSVAINATRPALHNASLPQAPTDGVLAAAVYQSEGEVHQSLERLESVITNT
                     SRVLKLLPDTIRRDHEQLLNLGLEAQMTELQQSRPTPQTQPRDTSSAKSSVFETYTLV
                     LIADSLSRNITKGVKRGVNEAIMLHLDHEVRHAIGNRLRQTQKNIIKSRLDEALKEST
                     TQFTAQLTQTVENLVKRELAEVLGSINGSLTSLVKENASLQKELNSIMSSGVLDEMRR
                     MREELCTLRESVAKRKATMPDSSLHATSSFQGRRSAPETILATALSMVREQQYRQGLE
                     YMLMAQQPSLLLRFLSILTRENENAYSELIENVETPNDVWCSVLLQLIEAAATEAEKE
                     VVVGVAIDILSERDQIAQNGALGSKLTTAMRAFERQARSETTSRSFLQCLKNLEKLLQ
                     S"
BASE COUNT      889 a    818 c    958 g    737 t
ORIGIN


  Query Match           100.0%;  Score 18;  DB 96;  Length 3402;
  Best Local Similarity  100.0%;  Pred. No. 2.3e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||||||||||||||||
Db   1459 TGCAGCAGCGGCAGAAGT 1442


RESULT   8
AR083116
LOCUS       AR083116    11873 bp    DNA             PAT       01-SEP-2000
DEFINITION  Sequence 32 from patent US 5976803.
ACCESSION   AR083116
VERSION     AR083116.1  GI:10009906
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 11873)
  AUTHORS   Meek,K.D.
  TITLE     Genetic test for equine severe combined immunodeficiency disease
  JOURNAL   Patent: US 5976803-A 32 02-NOV-1999;
FEATURES             Location/Qualifiers
     source          1. .11873
                     /organism="unknown"
BASE COUNT     3585 a    2320 c    2669 g    3299 t
ORIGIN


  Query Match           100.0%;  Score 18;  DB 9;  Length 11873;
  Best Local Similarity  100.0%;  Pred. No. 1.8e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||||||||||||||||
Db   6429 TGCAGCAGCGGCAGAAGT 6446


RESULT   9
AR083115
LOCUS       AR083115    11878 bp    DNA             PAT       01-SEP-2000
DEFINITION  Sequence 31 from patent US 5976803.
ACCESSION   AR083115
VERSION     AR083115.1  GI:10009905
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 11878)
  AUTHORS   Meek,K.D.
  TITLE     Genetic test for equine severe combined immunodeficiency disease
  JOURNAL   Patent: US 5976803-A 31 02-NOV-1999;
FEATURES             Location/Qualifiers
     source          1. .11878
                     /organism="unknown"
BASE COUNT     3586 a    2323 c    2668 g    3301 t
ORIGIN


  Query Match           100.0%;  Score 18;  DB 9;  Length 11878;
  Best Local Similarity  100.0%;  Pred. No. 1.8e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||||||||||||||||
Db   6429 TGCAGCAGCGGCAGAAGT 6446


RESULT  10
AR083114
LOCUS       AR083114    11883 bp    DNA             PAT       01-SEP-2000
DEFINITION  Sequence 28 from patent US 5976803.
ACCESSION   AR083114
VERSION     AR083114.1  GI:10009904
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 11883)
  AUTHORS   Meek,K.D.
  TITLE     Genetic test for equine severe combined immunodeficiency disease
  JOURNAL   Patent: US 5976803-A 28 02-NOV-1999;
FEATURES             Location/Qualifiers
     source          1. .11883
                     /organism="unknown"
BASE COUNT     3588 a    2324 c    2669 g    3302 t
```

```
ORIGIN


  Query Match             100.0%;  Score 18;  DB 9;  Length 11883;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaagt 18
          ||||||||||||||||||
Db   6432 TGCAGCAGCGGCAGAAGT 6449


RESULT  11
AC085350
LOCUS       AC085350      915 bp    DNA             HTG       30-NOV-2000
DEFINITION  Giardia intestinalis clone KJ5074 strain WB-C6, LOW-PASS SEQUENCE
            SAMPLING.
ACCESSION   AC085350
VERSION     AC085350.1  GI:11465342
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Giardia intestinalis.
  ORGANISM  Giardia intestinalis
            Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.
REFERENCE   1  (bases 1 to 915)
  AUTHORS   McArthur,A.G., Morrison,H.G., Nixon,J.E.J., Passamaneck,N.Q.E.,
            Kim,U., Hinkle,G., Crocker,M.K., Holder,M.E., Farr,R., Reich,C.I.,
            Olsen,G.J., Aley,S.B., Adam,R.D., Gillin,F.D. and Sogin,M.L.
  TITLE     The Giardia genome project database
  JOURNAL   FEMS Microbiol. Lett. 189 (2), 271-273 (2000)
  MEDLINE   20389616
REFERENCE   2  (bases 1 to 915)
  AUTHORS   Kim,U., Morrison,H.G., McArthur,A.G., Nixon,J., Eakin,N.Q.,
            Fierro,L.A., Aley,S.B. and Sogin,M.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-NOV-2000) Josephine Bay Paul Center for Comparative
            Molecular Biology and Evolution, Marine Biological Laboratory, 7
            MBL Street, Woods Hole, MA 02543-1015, USA
COMMENT     * NOTE: This record contains 1 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *       1      915: contig of 915 bp in length.
FEATURES            Location/Qualifiers
     source         1. .915
                    /organism="Giardia intestinalis"
                    /strain="WB-C6"
                    /db_xref="taxon:5741"
                    /clone="KJ5074"
BASE COUNT      278 a    195 c    274 g    166 t      2 others
ORIGIN


  Query Match             94.4%;  Score 17;  DB 77;  Length 915;
  Best Local Similarity   100.0%;  Pred. No. 8.5e+02;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaag 17
          |||||||||||||||||
Db    818 TGCAGCAGCGGCAGAAG 834


RESULT  12
AC034857/c
LOCUS       AC034857      920 bp    DNA             HTG       05-APR-2000
DEFINITION  Giardia intestinalis clone AJ1539 strain WB-C6, LOW-PASS SEQUENCE
            SAMPLING.
ACCESSION   AC034857
VERSION     AC034857.1  GI:7453065
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Giardia intestinalis.
  ORGANISM  Giardia intestinalis
            Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.
REFERENCE   1  (bases 1 to 920)
  AUTHORS   Morrison,H.G., McArthur,A.G., Nixon,J., Eakin,N.Q., Kim,U.,
            Crocker,M.C., Hinkle,G., Holder,M.E. and Sogin,M.L.
  TITLE     Giardia: a model for ancient eukaryotic genome analysis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 920)
  AUTHORS   McArthur,A.G., Morrison,H.G., Nixon,J., Eakin,N.Q., Kim,U.,
            Crocker,M.C., Hinkle,G., Holder,M.E. and Sogin,M.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-APR-2000) Josephine Bay Paul Center for Comparative
            Molecular Biology and Evolution, Marine Biological Laboratory, 7
            MBL Street, Woods Hole, MA 02543-1015, USA
COMMENT     * NOTE: This record contains 1 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *       1      920: contig of 920 bp in length.
FEATURES            Location/Qualifiers
     source         1. .920
                    /organism="Giardia intestinalis"
                    /strain="WB-C6"
                    /db_xref="taxon:5741"
                    /clone="AJ1539"
BASE COUNT      187 a    250 c    223 g    260 t
ORIGIN


  Query Match             94.4%;  Score 17;  DB 71;  Length 920;
  Best Local Similarity   100.0%;  Pred. No. 8.5e+02;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaag 17
          |||||||||||||||||
Db    678 TGCAGCAGCGGCAGAAG 662


RESULT  13
AC029282
LOCUS       AC029282     1009 bp    DNA             HTG       01-APR-2000
DEFINITION  Giardia intestinalis clone HF2663 strain WB-C6, LOW-PASS SEQUENCE
            SAMPLING.
ACCESSION   AC029282
VERSION     AC029282.1  GI:7384133
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Giardia intestinalis.
  ORGANISM  Giardia intestinalis
            Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.
REFERENCE   1  (bases 1 to 1009)
  AUTHORS   Morrison,H.G., McArthur,A.G., Nixon,J., Eakin,N.Q., Kim,U.,
            Crocker,M.C., Hinkle,G., Holder,M.E. and Sogin,M.L.
  TITLE     Giardia: a model for ancient eukaryotic genome analysis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1009)
  AUTHORS   Hinkle,G., Morrison,H.G., McArthur,A.G., Nixon,J., Eakin,N.Q.,
            Kim,U., Crocker,M.K., Holder,M.E. and Sogin,M.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-APR-2000) Josephine Bay Paul Center for Comparative
            Molecular Biology and Evolution, Marine Biological Laboratory, 7
            MBL Street, Woods Hole, MA 02543-1015, USA
COMMENT     * NOTE: This record contains 1 individual
```

```
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *       1     1009: contig of 1009 bp in length.
FEATURES             Location/Qualifiers
     source          1. .1009
                     /organism="Giardia intestinalis"
                     /strain="WB-C6"
                     /db_xref="taxon:5741"
                     /clone="HF2663"
BASE COUNT      299 a    211 c    290 g    207 t      2 others
ORIGIN


  Query Match             94.4%;  Score 17;  DB 71;  Length 1009;
  Best Local Similarity   100.0%;  Pred. No. 8.4e+02;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaag 17
          |||||||||||||||||
Db    766 TGCAGCAGCGGCAGAAG 782


RESULT  14
AF309358
LOCUS       AF309358     1671 bp    DNA             STS       30-MAR-2001
DEFINITION  Hordeum vulgare STS marker ABG452, sequence tagged site.
ACCESSION   AF309358
VERSION     AF309358.1  GI:13383483
KEYWORDS    STS.
SOURCE      barley.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae;
            Hordeum.
REFERENCE   1  (bases 1 to 1671)
  AUTHORS   Collins,N.C., Lahaye,T., Peterhansel,C., Freialdenhoven,A.,
            Corbitt,M. and Schulze-Lefert,P.
  TITLE     Sequence haplotypes revealed by sequence-tagged site fine mapping
            of the ror1 gene in the centromeric region of barley chromosome 1h
  JOURNAL   Plant Physiol. 125 (3), 1236-1247 (2001)
  MEDLINE   21140309
   PUBMED   11244105
REFERENCE   2  (bases 1 to 1671)
  AUTHORS   Collins,N.C., Lahaye,T., Peterhansel,C., Freialdenhoven,A.,
            Corbitt,M. and Schulze-Lefert,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-SEP-2000) Sainsbury Laboratory, John Innes Institute,
            Colney Lane, Norwich, Norfolk NR4 7UH, UK
COMMENT     primers: 5' segment: forward GGGAGTATGTGATATTGTGGGCATC, reverse
            ATAGTACAGACAGCCTCCTCTCATCC ; 3' segment: forward
            TCTACCTAGCTTCTTTCAAA, reverse AGCTGCCACCACCCCCAGTG.
FEATURES             Location/Qualifiers
     source          1. .1671
                     /organism="Hordeum vulgare"
                     /db_xref="taxon:4513"
                     /chromosome="1H"
                     /map="1H centromeric region"
                     /note="genotype: BCIngrid mlo-5"
     STS             <1. .1208
                     /standard_name="ABG452"
     primer_bind     1098. .1117
     STS             1098. .>1671
     primer_bind     complement(1183. .1208)
BASE COUNT      464 a    408 c    427 g    372 t
ORIGIN


  Query Match             94.4%;  Score 17;  DB 53;  Length 1671;
  Best Local Similarity   100.0%;  Pred. No. 7.7e+02;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaag 17
          |||||||||||||||||
Db    265 TGCAGCAGCGGCAGAAG 281


RESULT  15
AB030448S1/c
LOCUS       AB030448S1   4545 bp    DNA             ROD       14-APR-2000
DEFINITION  Mus musculus V1bR gene for V1b arginine vasopressin receptor, exon
            1.
ACCESSION   AB030448
VERSION     AB030448.1  GI:6683122
KEYWORDS    V1bR; V1b arginine vasopressin receptor.
SEGMENT     1 of 2
SOURCE      Mus musculus (strain:129SVJ) 4-8 weeks female tissue_lib:liver DNA.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Kikuchi,S., Tanoue,A., Goda,N., Matsuo,N. and Tsujimoto,G.
  TITLE     Structure and sequence of the mouse V1a and V1b vasopressin
            receptor genes
  JOURNAL   Jpn. J. Pharmacol. 81 (4), 388-392 (1999)
  MEDLINE   20132529
REFERENCE   2  (bases 1 to 4545)
  AUTHORS   Tsujimoto,G. and Kikuchi,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JUL-1999) to the DDBJ/EMBL/GenBank databases. Gozoh
            Tsujimoto, National Children's Medical Research Center, Molecular
            Pharmacology; 3-35-31 Taishido, Setagaya-ku, Tokyo-to 154-8509,
            Japan (E-mail:skikuchi@nch.go.jp, Tel:81-3-3414-8121(ex.2733),
            Fax:81-3-3419-1252)
FEATURES             Location/Qualifiers
     source          1. .4545
                     /organism="Mus musculus"
                     /strain="129SVJ"
                     /db_xref="taxon:10090"
                     /dev_stage="4-8 weeks"
                     /sex="female"
                     /tissue_lib="liver"
     exon            1424. .2354
                     /gene="V1bR"
                     /note="CDS is reported in Acc# AB030449"
                     /number=1
                     /product="V1b arginine vasopressin receptor"
BASE COUNT     1012 a   1246 c   1013 g   1274 t
ORIGIN


  Query Match             94.4%;  Score 17;  DB 94;  Length 4545;
  Best Local Similarity   100.0%;  Pred. No. 6.4e+02;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tgcagcagcggcagaag 17
          |||||||||||||||||
Db   2185 TGCAGCAGCGGCAGAAG 2169


Search completed: September 21, 2001, 00:35:06
Job time: 29561 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: September 20, 2001, 16:22:25 ; Search time 3479.37 Seconds
(without alignments)
2640.663 Million cell updates/sec

Title: US-09-138-735-1_COPY_1232_1825
Perfect score: 594
Sequence: 1 caggtacagcgtaacggctt.....aaacgagctccgtggcgata 594

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*
44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vi1:*
59: gb_vi2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	594	100.0	3402	9	AR047920	AR047920 Sequence
	2	590.8	99.5	3402	9	A48910	A48910 Sequence 1
	3	590.8	99.5	3402	96	TCU24190	U24190 Trypanosoma
	4	45	7.6	47852	3	MTV023	AL022022 Mycobacte
	5	44	7.4	2402	3	MLURHO	L27277 Micrococcus
c	6	43.2	7.3	2028	15	SSI132828	AJ132828 Spermatoz
c	7	43.2	7.3	158390	60	AC008532	AC008532 Homo sapi
c	8	42.8	7.2	767	6	EIMMAX	M30933 E.tenella a

c	9	42.6	7.2	140812	13	AP001081	AP001081 Oryza sat
	10	41.6	7.0	5880	59	HSE4	D14486 Equine herp
c	11	41.2	6.9	2479	95	RNU87960	U87960 Rattus norv
c	12	40.4	6.8	477	10	E34519	E34519 SCA7 gene a
c	13	40.4	6.8	477	88	AF020275	AF020275 Homo sapi
c	14	40.4	6.8	3932	93	HSM801297	AL133030 Homo sapi
c	15	40.4	6.8	3941	95	RNLAR2	X83546 R.norvegicu
c	16	40.4	6.8	5477	85	AB051453	AB051453 Homo sapi
	17	40.4	6.8	69350	3	MTV004	AL009198 Mycobacte
c	18	40.4	6.8	149618	91	AP000556	AP000556 Homo sapi
	19	40.4	6.8	150036	91	AP000557	AP000557 Homo sapi
c	20	40.4	6.8	157086	91	AP000552	AP000552 Homo sapi
c	21	40.4	6.8	166447	87	AC018751	AC018751 Homo sapi
	22	40.4	6.8	169237	87	AC009516	AC009516 Homo sapi
	23	40.4	6.8	171350	72	AC058816	AC058816 Homo sapi
	24	40.4	6.8	176051	88	AC023490	AC023490 Homo sapi
c	25	40.4	6.8	212656	86	AC007957	AC007957 Homo sapi
	26	40.2	6.8	701	13	AF310215	AF310215 Sorghum b
	27	40.2	6.8	138273	83	AP003297	AP003297 Oryza sat
c	28	40.2	6.8	172307	71	AC044842	AC044842 Homo sapi
	29	40	6.7	624	5	AF139019	AF139019 Cepaea ne
c	30	40	6.7	122395	78	AC090433	AC090433 Chlamydom
c	31	40	6.7	130569	78	AC090650	AC090650 Arabidops
	32	39.6	6.7	618	14	CRCGCR1	X17207 Chlamydomon
c	33	39.4	6.6	947	53	CNS0794Y	AL434840 T3 end of
	34	39.4	6.6	1442	88	AF087653	AF087653 Homo sapi
c	35	39.2	6.6	818	8	CHKPROTAMI	L38713 Gallus gall
c	36	39.2	6.6	1528	8	CHKPROTAMN	M28100 Gallus gall
	37	39.2	6.6	2771	94	MMZINCFPR	X95503 M.musculus
	38	39.2	6.6	3732	94	MMZINCFIP	X95504 M.musculus
	39	39.2	6.6	156589	92	HS537K23	AL034405 Human DNA
c	40	39	6.6	259	85	AB018491	AB018491 Homo sapi
c	41	38.8	6.5	720	85	AB007820	AB007820 Homo sapi
	42	38.8	6.5	1030	14	HVCAB2	X12735 Barley Cab-
c	43	38.8	6.5	2858	97	MFAPOA4A	X68361 M.fascicula
c	44	38.6	6.5	54327	78	AC090435	AC090435 Chlamydom
c	45	38.6	6.5	88050	78	AC090436	AC090436 Chlamydom

ALIGNMENTS

```
RESULT   1
AR047920
LOCUS       AR047920      3402 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 1 from patent US 5820864.
ACCESSION   AR047920
VERSION     AR047920.1  GI:5970263
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 3402)
  AUTHORS   Paranhos-Baccala,G., Lesenechal,M. and Jolivet,M.
  TITLE     Trypanosoma cruzi antigen, gene encoding therefor and methods of
            detecting and treating chagas disease
  JOURNAL   Patent: US 5820864-A 1 13-OCT-1998;
FEATURES             Location/Qualifiers
     source          1. .3402
                     /organism="unknown"
BASE COUNT      888 a    821 c    956 g    737 t
ORIGIN


  Query Match             100.0%;  Score 594;  DB 9;  Length 3402;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-114;
  Matches  594;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 caggtacagcgtaacggcttttgcttcaatcgtacagccgacggtagctgcgtcctggct 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1232 CAGGTACAGCGTAACGGCTTTTGCTTCAATCGTACAGCCGACGGTAGCTGCGTCCTGGCT 1291

Qy     61 gacatgtcgattcgattgacgatcttccatctccggtcctcccgcagggaagaacagcag 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1292 GACATGTCGATTCGATTGACGATCTTCCATCTCCGGTCCTCCCGCAGGGAAGAACAGCAG 1351

Qy    121 ccaggccaaaaaaacatcggtagtggcgacggcgaaaccggggtgtgtgtcctcgggcact 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1352 CCAGGCCAAAAAACATCGGTAGTGGCGACGGCGAAACCGGGGTGTGTGTCCTCGGGCACT 1411

Qy    181 gacgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatcccctgcatca 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1412 GACGCGGCGAGTAGCAGTCATACCAATACGACTTCTGCCGCTGCTGCATCCCCTGCATCA 1471

Qy    241 ccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggctgag 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1472 CCCCCTGTTTCAGCGCCAGCCAAGGCAGCCGCGCCTCCTGCCGCGGCGCGATCGGCTGAG 1531

Qy    301 ccgcacgtggggagcaagatcattgctaatctagtgaatcagctggggattaatgtcacc 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1532 CCGCACGTGGGGAGCAAGATCATTGCTAATCTAGTGAATCAGCTGGGGATTAATGTCACC 1591

Qy    361 caaaggagcgtcgtcagcactggagcgccggccacgacgaggtctacggcggtgacgtcc 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1592 CAAAGGAGCGTCGTCAGCACTGGAGCGCCGGCCACGACGAGGTCTACGGCGGTGACGTCC 1651

Qy    421 acgactaccgccccgcagcgaacaagtccatacgggcacaatggccgacctgtgacggct 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1652 ACGACTACCGCCCCGCAGCGAACAAGTCCATACGGGCACAATGGCCGACCTGTGACGGCT 1711

Qy    481 ggattggtggcagctaatagtggtgccagcgcggcctcgtctcccacagccgcggcgaaa 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1712 GGATTGGTGGCAGCTAATAGTGGTGCCAGCGCGGCCTCGTCTCCCACAGCCGCGGCGAAA 1771

Qy    541 ccaacaggagaagaaaaggcctccgcggcatgtgaaacgagctccgtggcgata 594
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1772 CCAACAGGAGAAGAAAAGGCCTCCGCGGCATGTGAAACGAGCTCCGTGGCGATA 1825


RESULT   2
A48910
LOCUS       A48910       3402 bp    DNA             PAT       07-MAR-1997
DEFINITION  Sequence 1 from Patent WO9605312.
ACCESSION   A48910
VERSION     A48910.1  GI:2302570
KEYWORDS    .
SOURCE      Trypanosoma cruzi.
  ORGANISM  Trypanosoma cruzi
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma; Schizotrypanum.
REFERENCE   1  (bases 1 to 3402)
  AUTHORS   Paranhos-Baccala,G., Lesenechal,M. and Jolivet,M.
  TITLE     NOVEL TRYPANOSOMA CRUZI ANTIGEN, AND GENE CODING THEREFOR; THEIR
            APPLICATION TO THE DETECTION OF CHAGAS' DISEASE
  JOURNAL   Patent: WO 9605312-A 1 22-FEB-1996;
            BIO MERIEUX (FR)
COMMENT     Other publication CA 2173957 960222
            Other publication AU 3169195 960307
            Other publication FR 2723589 960216.
FEATURES             Location/Qualifiers
     source          1. .3402
                     /organism="Trypanosoma cruzi"
                     /strain="G"
                     /db_xref="taxon:5693"
                     /dev_stage="EPIMASTIGOTE"
BASE COUNT      889 a    818 c    958 g    737 t
ORIGIN


  Query Match             99.5%;  Score 590.8;  DB 9;  Length 3402;
  Best Local Similarity   99.7%;  Pred. No. 6.9e-114;
  Matches  592;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 caggtacagcgtaacggcttttgcttcaatcgtacagccgacggtagctgcgtcctggct 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1232 CAGGTACAGCGTAACGGCTTTTGCTTCAATCGTACAGCCGACGGTAGCTGCGTCCTGGCT 1291
```

```
Qy      61 gacatgtcgattcgattgacgatcttccatctccggtcctcccgcagggaagaacagcag 120
           |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1292 GACATGTCGAATCGATTGACGATCTTCCATCTCCGGTCCTCCCGCAGGGAAGAACAGCAG 1351

Qy     121 ccaggccaaaaaacatcggtagtggcgacggcgaaaccggggtgtgtgtcctcgggcact 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1352 CCAGGCCAAAAAACATCGGTAGTGGCGACGGCGAAACCGGGGTGTGTGTCCTCGGGCACT 1411

Qy     181 gacgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatcccctgcatca 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1412 GACGCGGCGAGTAGCAGTCATACCAATACGACTTCTGCCGCTGCTGCATCCCCTGCATCA 1471

Qy     241 ccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggctgag 300
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    1472 CCCCCTGTTTCAGCGCCAGCGAAGGCAGCCGCGCCTCCTGCCGCGGCGCGATCGGCTGAG 1531

Qy     301 ccgcacgtggggagcaagatcattgctaatctagtgaatcagctggggattaatgtcacc 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1532 CCGCACGTGGGGAGCAAGATCATTGCTAATCTAGTGAATCAGCTGGGGATTAATGTCACC 1591

Qy     361 caaaggagcgtcgtcagcactggagcgccggccacgacgaggtctacggcggtgacgtcc 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1592 CAAAGGAGCGTCGTCAGCACTGGAGCGCCGGCCACGACGAGGTCTACGGCGGTGACGTCC 1651

Qy     421 acgactaccgccccgcagcgaacaagtccatacgggcacaatggccgacctgtgacggct 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1652 ACGACTACCGCCCCGCAGCGAACAAGTCCATACGGGCACAATGGCCGACCTGTGACGGCT 1711

Qy     481 ggattggtggcagctaatagtggtgccagcgcggcctcgtctcccacagccgcggcgaaa 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1712 GGATTGGTGGCAGCTAATAGTGGTGCCAGCGCGGCCTCGTCTCCCACAGCCGCGGCGAAA 1771

Qy     541 ccaacaggagaagaaaaggcctccgcggcatgtgaaacgagctccgtggcgata 594
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1772 CCAACAGGAGAAGAAAAGGCCTCCGCGGCATGTGAAACGAGCTCCGTGGCGATA 1825
```

```
RESULT   3
TCU24190
LOCUS       TCU24190     3402 bp    mRNA            INV       04-AUG-1997
DEFINITION  Trypanosoma cruzi Tc40 antigen (Tc40) mRNA, complete cds.
ACCESSION   U24190
VERSION     U24190.1  GI:790645
KEYWORDS    .
SOURCE      Trypanosoma cruzi.
  ORGANISM  Trypanosoma cruzi
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma; Schizotrypanum.
REFERENCE   1  (bases 1 to 3402)
  AUTHORS   Lesenechal,M., Duret,L., Cano,M.I., Mortara,R.A., Jolivet,M.,
            Camargo,M.E., da Silveira,J.F. and Paranhos-Baccala,G.
  TITLE     Cloning and characterization of a gene encoding a novel
            immunodominant antigen of Trypanosoma cruzi
  JOURNAL   Mol. Biochem. Parasitol. 87 (2), 193-204 (1997)
  MEDLINE   97391123
REFERENCE   2  (bases 1 to 3402)
  AUTHORS   Lesenechal,M., Franco Da Silveira,J., Mortara,R.A., Duret,L.,
            Camargo,M.E., Jolivet,M. and Paranhos-Baccala,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-APR-1995) Mylene Lesenechal, BioMerieux, 22 rue Saint
            Jean de Dieu, Lyon 69007, France
FEATURES             Location/Qualifiers
     source          1. .3402
                     /organism="Trypanosoma cruzi"
                     /strain="G"
                     /db_xref="taxon:5693"
                     /dev_stage="epimastigote"
     misc_RNA        1. .35
                     /gene="Tc40"
                     /note="spliced leader"
     gene            1. .3402
                     /gene="Tc40"
     CDS             266. .3013
                     /gene="Tc40"
                     /codon_start=1
                     /product="Tc40 antigen"
                     /protein_id="AAC47657.1"
                     /db_xref="GI:790646"
                     /translation="MTVTVDLFNHAKPSNNEGRVWSVDAATFNEVPEAQRVLADSQFY
                     LAYTMKRRHVLRVVKRSNLLKGTVRAHSKPIHAVKFVNYRSNVAASAGKGEFFVWVVT
                     DETEASNGKPDLAARLTVKVYFKLQDPVTIPCFSFFINAESQRPDLLVLYETQAAILD
                     SSSLIERFDVESLEATLQRNCTTLRTLTQPVSENSLCSVGSGGWFTFTTEPTMVAACT
                     LRNRSTPSWACCEGEPVKALHLLDATVEENVSVLVAASTKGVYQWLLTGVAEPNLLRK
                     FVIDGSIVAMESSRETFAVFDDRKQLALVNMHSPHNFTCTHYMMPCQVQRNGFCFNRT
                     ADGSCVLADMSNRLTIFHLRSSRREEQQPGQKTSVVATAKPGCVSSGTDAASSSHTNT
                     TSAAAASPASPPVSAPAKAAAPPAAARSAEPHVGSKIIANLVNQLGINVTQRSVVSTG
                     APATTRSTAVTSTTTAPQRTSPYGHNGRPVTAGLVAANSGASAASSPTAAAKPTGEEK
                     ASAACETSSVAINATRPALHNASLPQAPTDGVLAAAVYQSEGEVHQSLERLESVITNT
                     SRVLKLLPDTIRRDHEQLLNLGLEAQMTELQQSRPTPQTQPRDTSSAKSSVFETYTLV
                     LIADSLSRNITKGVKRGVNEAIMLHLDHEVRHAIGNRLRQTQKNIIKSRLDEALKEST
                     TQFTAQLTQTVENLVKRELAEVLGSINGSLTSLVKENASLQKELNSIMSSGVLDEMRR
                     MREELCTLRESVAKRKATMPDSSLHATSSFQGRRSAPETILATALSMVREQQYRQGLE
                     YMLMAQQPSLLLRFLSILTRENENAYSELIENVETPNDVWCSVLLQLIEAAATEAEKE
                     VVVGVAIDILSERDQIAQNGALGSKLTTAMRAFERQARSETTSRSFLQCLKNLEKLLQ
                     S"
BASE COUNT      889 a    818 c    958 g    737 t
ORIGIN
```

```
  Query Match             99.5%;  Score 590.8;  DB 96;  Length 3402;
  Best Local Similarity   99.7%;  Pred. No. 6.9e-114;
  Matches  592;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 caggtacagcgtaacggcttttgcttcaatcgtacagccgacggtagctgcgtcctggct 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1232 CAGGTACAGCGTAACGGCTTTTGCTTCAATCGTACAGCCGACGGTAGCTGCGTCCTGGCT 1291

Qy      61 gacatgtcgattcgattgacgatcttccatctccggtcctcccgcagggaagaacagcag 120
           |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1292 GACATGTCGAATCGATTGACGATCTTCCATCTCCGGTCCTCCCGCAGGGAAGAACAGCAG 1351

Qy     121 ccaggccaaaaaacatcggtagtggcgacggcgaaaccggggtgtgtgtcctcgggcact 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1352 CCAGGCCAAAAAACATCGGTAGTGGCGACGGCGAAACCGGGGTGTGTGTCCTCGGGCACT 1411

Qy     181 gacgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatcccctgcatca 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1412 GACGCGGCGAGTAGCAGTCATACCAATACGACTTCTGCCGCTGCTGCATCCCCTGCATCA 1471

Qy     241 ccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggctgag 300
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    1472 CCCCCTGTTTCAGCGCCAGCGAAGGCAGCCGCGCCTCCTGCCGCGGCGCGATCGGCTGAG 1531

Qy     301 ccgcacgtggggagcaagatcattgctaatctagtgaatcagctggggattaatgtcacc 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1532 CCGCACGTGGGGAGCAAGATCATTGCTAATCTAGTGAATCAGCTGGGGATTAATGTCACC 1591

Qy     361 caaaggagcgtcgtcagcactggagcgccggccacgacgaggtctacggcggtgacgtcc 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1592 CAAAGGAGCGTCGTCAGCACTGGAGCGCCGGCCACGACGAGGTCTACGGCGGTGACGTCC 1651

Qy     421 acgactaccgccccgcagcgaacaagtccatacgggcacaatggccgacctgtgacggct 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1652 ACGACTACCGCCCCGCAGCGAACAAGTCCATACGGGCACAATGGCCGACCTGTGACGGCT 1711

Qy     481 ggattggtggcagctaatagtggtgccagcgcggcctcgtctcccacagccgcggcgaaa 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1712 GGATTGGTGGCAGCTAATAGTGGTGCCAGCGCGGCCTCGTCTCCCACAGCCGCGGCGAAA 1771

Qy     541 ccaacaggagaagaaaaggcctccgcggcatgtgaaacgagctccgtggcgata 594
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1772 CCAACAGGAGAAGAAAAGGCCTCCGCGGCATGTGAAACGAGCTCCGTGGCGATA 1825
```

```
RESULT   4
```

```
MTV023
LOCUS       MTV023      47852 bp    DNA             BCT       17-JUN-1998
DEFINITION  Mycobacterium tuberculosis H37Rv complete genome; segment 148/162.
ACCESSION   AL022022 AL123456
VERSION     AL022022.1  GI:3261554
KEYWORDS    .
SOURCE      Mycobacterium tuberculosis.
  ORGANISM  Mycobacterium tuberculosis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE   1  (bases 1 to 47852)
  AUTHORS   Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
            Harris,D., Gordon,S.V., Eiglmeier,K., Gas,S., Barry III,C.E.,
            Tekaia,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
            Connor,R., Davies,R., Devlin,K., Feltwell,T., Gentles,S.,
            Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J.,
            Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
            Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S.,
            Squares,S., Sqares,R., Sulston,J.E., Taylor,K., Whitehead,S. and
            Barrell,B.G.
  TITLE     Deciphering the biology of Mycobacterium tuberculosis from the
            complete genome sequence
  JOURNAL   Nature 393 (6685), 537-544 (1998)
  MEDLINE   98295987
  REMARK    Erratum:[[published erratum appears in Nature 1998 Nov
            12;396(6707):190]]
REFERENCE   2  (bases 1 to 47852)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
            tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
            Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
            Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
            75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
COMMENT     On Jun 27, 1998 this sequence version replaced gi:2924430.
            Notes:
            Details of M. tuberculosis sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have
            been renumbered from the original cosmid submissions but the old
            gene designations are in brackets after the new gene numbers.
            Gene prediction was based on a Hidden Markov Model of TB genes
            implemented in TBparse (Krogh) supplemented with visual inspection
            of positional base preference in codons, especially where there is
            an increase in the observed/expected third position G + C.
            CAUTION:  In some cases we may not have predicted the correct
            initiation codon.  Where possible we choose an initiation codon
            (atg, gtg, or ttg) which is preceded by an upstream ribosome
            binding site sequence (optimally 5-13bp before the initiation
            codon).  If this cannot be identified we choose the most upstream
            initiation codon.
FEATURES             Location/Qualifiers
     source          1. .47852
                     /organism="Mycobacterium tuberculosis"
                     /strain="H37Rv"
                     /db_xref="taxon:1773"
     source          <1. .18
                     /organism="Mycobacterium tuberculosis"
                     /strain="H37Rv"
                     /db_xref="taxon:1773"
                     /clone="Y13E12"
     misc_feature    <1. .47852
                     /note="fragment designated v023. Does not represent a
                     physical clone"
     gene            complement(1. .1695)
                     /gene="Rv3494c"
     CDS             complement(1. .1695)
                     /gene="Rv3494c"
                     /note="Rv3494c, (MTV023.01c), len: 564. Unknown Pro-rich
                     protein similar to several Mycobacterium tuberculosis
                     proteins e.g. MTCI28.14 (515 aa), MTCY19H5.28c (516 aa)and
                     (MTV051.09). Has hydrophobic stretch, possibly signal
                     peptide at N-terminus. FASTA scores: Z97050|MTCI28_14
                     (515 aa)  opt: 1006 z-score: 582.8 E(): 5.9e-25;    34.3%
                     identity in 559 aa overlap; Z97182|MTCY19H5_28  (516 aa)
                     opt:  979 z-score:567.7 E(): 4.1e-24; 33.5% identity in
                     555 aa overlap. TBparse score is 0.897"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein Rv3494c"
                     /protein_id="CAA17731.1"
                     /db_xref="GI:2924431"
                     /db_xref="SPTREMBL:O53539"
                     /translation="MIDRLAKIQLSIFAVITVITLSVMAIFYLRLPATFGIGTYGVSA
                     DFVAGGGLYKNANVTYRGVAVGRVESVGLNPNGVTAHMRLNSGTAIPSNVTATVRSVS
                     AIGEQYIDLVPPENPSSTKLRNGFRIQRQNTRIGQDVADLLRQAETLLGSLGDTRLRE
                     LLHEAFIATNGAGPELARLIESARLLVDEANANYPQVSQLIDQAGPFLQAQIRAGGDI
                     KSLADGLARFTWQLRAADPRLRDTLADAPDAIDEANTAFSGIRPSFPALAASLANLGR
                     VGVIYHKSIEQLLVVFPALFAAIITSAGGVPQDEGAKLDFKIDLHDPPPCMTGFLPPP
                     LVRSPADESVREIPRDMYCKTAQNDPSTVRGARNYPCQEFPGKRAPTVQLCRDPRGYV
                     PVGTNPWRGPPIPYGTEVTDGRNILPPNKFPYIPPGADPDPGVPIVGPPPPGQVAGPG
                     PAPHQPAQPAPPPNDNGPPPPFTSWMPPGYPPEPPQVPYPATIPPPPPPEGTGPPPGP
                     APGPQPQASGPAYTIYDQLSGAFADPAGGTGIFAPGMTGASSAENWVDLMRDPRQL"
     RBS             complement(1703. .1707)
                     /note="possible RBS for Rv3494c"
     gene            complement(1706. .2860)
                     /gene="lprN"
     CDS             complement(1706. .2860)
                     /gene="lprN"
                     /note="Rv3495c, (MTV023.02c), len: 384. lprN,  similar to
                     Mycobacterium tuberculosis proteins MTCI28.13 (390 aa) and
                     MTCY19H5.29 (402 aa) and (MTV051.08). Probably
                     lipoprotein,contains possible signal sequence and
                     appropriately positioned PS00013 Prokaryotic membrane
                     lipoprotein lipid attachment site. FASTA scores:
                     gp|Z97050|MTCI28_13 (390 aa)  opt:653 z-score: 762.1 E():
                     0;    33.6% identity in 363 aa overlap; and
                     Z97182|MTCY19H5_29 (402 aa)  opt: 572 z-score: 667.9 E():
                     1.1e-29;    31.8% identity in 362 aa overlap. TBparse
                     score is 0.897"
                     /codon_start=1
                     /transl_table=11
                     /product="lprN"
                     /protein_id="CAA17732.1"
                     /db_xref="GI:2924432"
                     /db_xref="SPTREMBL:O53540"
                     /translation="MNRIWLRAIILTASSALLAGCQFGGLNSLPLPGTAGHGEGAYSV
                     TVEMADVATLPQNSPVMVDDVTVGSVAGIVAVQRPDGSFYAAVKLDLDKNVLLPANAV
                     AKVSQTSLLGSLHVELAPPTDRPPTGRLVDGSRITEANTDRFPTTEEVFSALGVVVNK
                     GNVGALEEIIDETHQAVAGRQAQFVNLVPRLAELTAGLNRQVHDIIDALDGLNRVSAI
                     LARDKDNLGRALDTLPDAVRVLNQNRDHIVDAFAALKRLTMVTSHVLAETKVDFGEDL
                     KDLYSIVKALNDDRKDFVTSLQLLLTFPFPNFGIKQAVRGDYLNVFTTFDLTLRRIGE
                     TFFTTAYFDPNMAHMDEILNPPDFLIGELANLSGQAADPFKIPPGTASGQ"
     misc_feature    complement(2798. .2830)
                     /gene="lprN"
                     /note="PS00013 Prokaryotic membrane lipoprotein lipid
                     attachment site"
     gene            complement(2857. .4212)
                     /gene="Rv3496c"
     CDS             complement(2857. .4212)
                     /gene="Rv3496c"
                     /note="Rv3496c, (MTV023.03c), len: 451. Unknown but
                     similar to Mycobacterium tuberculosis proteins MTCI28.12
                     (530 aa), MTCY19H5.30c (508 aa) (MTV051.07). Hydrophobic
                     region at N-terminus. FASTA scores: Z97050|MTCI28_12  (530
                     aa) opt:838 z-score: 711.0 E(): 4.3e-32;    35.1% identity
                     in 473 aaoverlap; and Z97182|MTCY19H5_30 (508 aa) opt:
                     821 z-score:697.0 E(): 2.6e-31; 35.1% identity in 453 aa
                     overlap.TBparse score is 0.891"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein Rv3496c"
                     /protein_id="CAA17733.1"
                     /db_xref="GI:2924433"
                     /db_xref="SPTREMBL:O53541"
```

```
                /translation="MMGRVAMLTGSRGLRYATVIALVAALVGGVYVLSSTGNKRTIVG
                YFTSAVGLYPGDQVRVLGVPVGEIDMIEPRSSDVKITMSVSKDVKVPVDVQAVIMSPN
                LVAARFIQLTPVYTGGAVLPDNGRIDLDRTAVPVEWDEVKEGLTRLAADLSPAAGELQ
                GPLGAAINQAADTLDGNGDSLHNALRELAQVAGRLGDSRGDIFGTVKNLQVLVDALSE
                SDEQIVQFAGHVASVSQVLADSSANLDQTLGTLNQALSDIRGFLRENNSTLIETVNQL
                NDFAQTLSDQSENIEQVLHVAGPGITNFYNIYDPAQGTLNGLLSIPNFANPVQFICGG
                SFDTAAGPSAPDYYRRAEICRERLGPVLRRLTVNYPPIMFHPLNTITAYKGQIIYDTP
                ATEAKSETPVPELTWVPAGGGAPVGNPADLQSLLVPPAPGPAPAPPAPGAGPGEHGGG
                G"
     RBS        complement(2867. .2871)
                /gene="Rv3496c"
                /note="possible RBS for Rv3495c"
     gene       complement(4209. .5282)
                /gene="Rv3497c"
     CDS        complement(4209. .5282)
                /gene="Rv3497c"
                /note="Rv3497c, (MTV023.04c), len: 357. Unknown but
                similar to Mycobacterium tuberculosis proteins MTCY19H5.31
                (481 aa), MTCI28.11, (515 aa)and MTV051.06. Hydrophobic
                region atN-terminus. FASTA scores: Z97182|MTCY19H5_31
                (481 aa) opt:611 z-score: 667.7 E(): 1.1e-29;    32.5%
                identity in 332aa overlap; and Z97050|MTCI28_11 (515 aa)
                opt:  587 z-score:641.2 E(): 3.3e-28;  30.1% identity in
                335 aa overlap. TBparse score is 0.889"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein Rv3497c"
                /protein_id="CAA17734.1"
                /db_xref="GI:2924434"
                /db_xref="SPTREMBL:O53542"
                /translation="MLNRKPSSKHERDPLRTGIFGLVLVICVVLIAFGYSGLPFWPQG
                KTYDAYFTDAGGITPGNSVYVSGLKVGAVSAVSLAGNSAKVTFSVDRSIVVGDQSLAA
                IRTDTILGERSIAVSPAGSGKSTTIPLSRTTTPYTLNGVLQDLGRNANDLNRPQFEQA
                LNVFTQALHDATPQVRGAVDGLTSLSRALNRRDEALQGLLAHAKSVTSVLSERAEQVN
                KLVEDGNQLFAALDARRAALSALISGIDDVAAQISGFVADNRKEFGPALSKLNLVLAN
                LNERRDYITEALKRLPTYATTLGEVVGSGPGFNVNVYSVLPGPLVATVFDLVFQPGKL
                PDSLADYLRGFIQERWIIRPKSP"
     gene       complement(5272. .6324)
                /gene="Rv3498c"
     CDS        complement(5272. .6324)
                /gene="Rv3498c"
                /note="Rv3498c, (MTV023.05c), len: 350. Unknown but
                similar to Mycobacterium tuberculosis proteins  MTCI28.10
                (346 aa), MTCY19H5.32c (346 aa) and MTV051.05. Hydrophobic
                region at N-terminus. FASTA scores: Z97050|MTCI28_10  (346
                aa)  opt: 755 z-score: 883.6 E():  0;    36.8% identity in
                340 aaoverlap; and Z97182|MTCY19H5_32 Mycobacterium
                tuberculosiscosmi (275 aa) opt:  699z-score: 820.1 E(): 0;
                47.0% identity in 249 aa overlap. TBparse score is 0.878"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein Rv3498c"
                /protein_id="CAA17735.1"
                /db_xref="GI:2924435"
                /db_xref="SPTREMBL:O53543"

  Query Match             7.6%;  Score 45;  DB 3;  Length 47852;
  Best Local Similarity   43.7%;  Pred. No. 2.8;
  Matches  198;  Conservative    0;  Mismatches  255;  Indels    0;  Gaps    0;

Qy     118 cagccaggccaaaaaacatcggtagtggcgacggcgaaaccggggtgtgtgtcctcgggc 177
           | | | ||| |     |  |||  | |  | |||| |     |    |  | | | || ||| |
Db   32502 CCGGCGGGCAAGGCGGCGCCGGCGGCGATGGCGGCAACGGGGCCATCGGCGGCCACGGCC 32561

Qy     178 actgacgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatcccctgca 237
             | |||          ||     |  |  | ||  | || ||| |  | |||
Db   32562 CACTCACTGACGACCCCGGCGGCAACGGGGGCACCGGCGGCAACGGCGGCACCGGCGGCA 32621

Qy     238 tcaccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggct 297
             |  |  |     || || |||||  ||| || || |   | ||   ||||  | ||||
Db   32622 CCGGCGGCGCGGGCATCGGCAGCCTTGGCGGCGGCACTGGCGGCGATGGCGGCAACGGCG 32681

Qy     298 gagccgcacgtggggagcaagatcattgctaatctagtgaatcagctggggattaatgtc 357
           |   ||   ||   |     ||     || |     | |   | |  ||        ||
Db   32682 GCAACGGCGGTACCGGCGGCGAGGGCGGCGAGGTCGGCGGCGCCGGCGGCACCGGCGGTG 32741

Qy     358 acccaaaggagcgtcgtcagcactggagcgccggccacgacgaggtctacggcggtgacg 417
             |     |  |  |   | | | || |||| || ||  |   |||| || | ||
Db   32742 CGGCCGGCAATGGCGGCGATGGCGGCACCGGCGGCACCGGCGGCGGGGACGGGGGCGCCG 32801

Qy     418 tccacgactaccgccccgcagcgaacaagtccatacgggcacaatggccgacctgtgacg 477
           |  |  |  | || |||  |  | |     | |  | | |   ||     | |  ||
Db   32802 GCGGCACCGGCGGCACCGGCGGCACCGGCGGCCTCGGCGACCCCCGGGTCGGCGGATCCG 32861

Qy     478 gctggattggtggcagctaatagtggtgccagcgcggcctcgtctcccacagccgcggcg 537
           || |    || |||| |       |  ||      | |  |     |  || | |||||
Db   32862 GCGGCGACGGCGGCACCGGCGGCAGCGGCGGTGCGGCCGGCAATGGCGGCAACGGCGGCA 32921

Qy     538 aaaccaacaggagaagaaaaggcctccgcggca 570
           |  ||  |  | || | || |||  | ||||||
Db   32922 ACGCCGGCGCGGGAGGCAATGGCAACGGCGGCA 32954


RESULT   5
MLURHO
LOCUS       MLURHO       2402 bp    DNA             BCT       12-NOV-1996
DEFINITION  Micrococcus luteus Rho factor (rho) gene, complete cds.
ACCESSION   L27277
VERSION     L27277.1  GI:1666539
KEYWORDS    .
SOURCE      Micrococcus luteus.
  ORGANISM  Micrococcus luteus
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Micrococcineae; Micrococcaceae; Micrococcus.
REFERENCE   1  (bases 905 to 2402)
  AUTHORS   Opperman,T. and Richardson,J.P.
  TITLE     Phylogenetic analysis of sequences from diverse bacteria with
            homology to the Escherichia coli rho gene
  JOURNAL   J. Bacteriol. 176 (16), 5033-5043 (1994)
  MEDLINE   94327472
REFERENCE   2  (bases 1 to 2402)
  AUTHORS   Nowatzke,W.L. and Richardson,J.P.
  TITLE     Characterization of an unusual Rho factor from the high G + C
            gram-positive bacterium Micrococcus luteus
  JOURNAL   J. Biol. Chem. 271 (2), 742-747 (1996)
  MEDLINE   96132802
REFERENCE   3  (bases 905 to 2402)
  AUTHORS   Richardson,J.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-FEB-1994) John P. Richardson, Chemistry, Indiana
            University, Bloomington, IN 47405, USA
REFERENCE   4  (bases 1 to 2402)
  AUTHORS   Nowatzke,W.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-NOV-1996) William L. Nowatzke, Chemistry, Indiana
            University, Bloomington, IN 47405, USA
COMMENT     On Nov 12, 1996 this sequence version replaced gi:968908.
FEATURES             Location/Qualifiers
     source          1. .2402
                     /organism="Micrococcus luteus"
                     /strain="EM"
                     /db_xref="taxon:1270"
     gene            288. .2363
                     /gene="rho"
     CDS             288. .2363
                     /gene="rho"
                     /function="transcription termination factor"
                     /codon_start=1
                     /transl_table=11
                     /product="Rho factor"
                     /protein_id="AAB18671.1"
                     /db_xref="GI:1666540"
                     /translation="MTESTEQTTPTNGGGLASLKLAQLQALASQLGIAGGSRMRKADL
                     VTAISDHQRGGSVADRDAAERAAQAPAAPAAETAPAAASSEDAAPAAERPARRRSRRA
                     DADTSAPAAAQDGQPQAEAREAQTEQAPRETASDQDRSGGSEARDEGEDRPQSERRSR
                     GRRRAGDDDAQQGQDRRSDGAQGEDGADADRRGDREDRDDNGRENGRGRNGRNGRDRD
```

NGRDRENGRENSRDRENGRDGSREQRGDKSEDGGRGDGGRGDRSRRDDRDDEGGRNRR
NRRNRNERGRNRRGRGGPEVDETELTEDDVLQPVAGILDVLDNYAFVRTSGYLPGPND
VYVSLAMVKKYGLRKGDAVVGPIAPRDGEKQQHHGGGSNRQKFNALVKISSVNGQPAV
EHPQRVEFGKLVPLYPQERLRLETDPKLIGPRVIDLVSPIGKGQRGLIVSPPKAGKTM
ILQSIANAIKTNNPEVHLMMVLVDERPEEVTDMQRSVDGEVIASTFDRPADDHTTLAE
LAIERAKRLVEMGRDVVVLLDSMTRLGRAYNLAAPASGRILSGGVDSSALYPPKKFFG
AARNIENGGSLTILATALVETGSRMDEVIFEEFKGTGNMELRLSRHLAERRIFPAVDV
NASGTRREEALLSQEEVKIMWKLRRVLSGLEQQQALDLLTNKIKDTASNAEFLMLVSK
TTLGSKGDD"

BASE COUNT 386 a 947 c 818 g 251 t
ORIGIN

Query Match 7.4%; Score 44; DB 3; Length 2402;
Best Local Similarity 49.6%; Pred. No. 12;
Matches 113; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

```
Qy    367 agcgtcgtcagcactggagcgccggccacgacgaggtctacggcggtgacgtccacgact 426
Db    799 AGCAGGGTCAGGACCGTCGCTCCGACGGCGCCCAGGGCGAGGACGGCGCCGACGCGGACC 858

Qy    427 accgccccgcagcgaacaagtccatacgggcacaatggccgacctgtgacggctggattg 486
Db    859 GCCGCGGCGACCGGGAGGACCGTGACGACAACGGGCGCGAGAACGGCCGCGGCCGCAACG 918

Qy    487 gtggcagctaatagtggtgccagcgcggcctcgtctcccacagccgcggcgaaaccaaca 546
Db    919 GCCGCAACGGGCGGGACCGCGACAACGGCCGCGACCGCGAGAACGGCCGCGAGAACAGCC 978

Qy    547 ggagaagaaaaggcctccgcggcatgtgaaacgagctccgtggcgata 594
Db    979 GGGACCGTGAGAACGGCCGCGACGGCTCGCGCGAGCAGCGCGGCGACA 1026
```

RESULT 6
SSI132828/c
LOCUS SSI132828 2028 bp mRNA PLN 05-APR-1999
DEFINITION Spermatozopsis similis mRNA for p210 protein, partial.
ACCESSION AJ132828
VERSION AJ132828.1 GI:4584085
KEYWORDS p210 protein.
SOURCE Spermatozopsis similis.
ORGANISM Spermatozopsis similis
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;
Spermatozopsis.
REFERENCE 1 (bases 1 to 2028)
AUTHORS Lechtreck,K.F., Teltenkoetter,A. and Grunow,A.
TITLE A 210-kDa-protein is located in a membrane-microtubule-linker at the distal end of mature and nascent basal bodies
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2028)
AUTHORS Lechtreck,K.F.
TITLE Direct Submission
JOURNAL Submitted (17-FEB-1999) Lechtreck K.F., Botany, University of Cologne, Gyrhofstr. 15, D-50931 Cologne, GERMANY
FEATURES Location/Qualifiers
source 1. .2028
/organism="Spermatozopsis similis"
/db_xref="taxon:3192"
CDS <1. .1830
/function="putative microtubule-membrane-linker"
/codon_start=1
/product="p210 protein"
/protein_id="CAB40552.1"
/db_xref="GI:4584086"
/translation="APAHAPPPLPSPQVLAQQAARALSMAPPPPPSALQQAASQPAPP
AYAPPPQPASYAPQPAQQQAPAPPTYAPPPAQQQAPQAPAYAPPAPAQQQAPAPQAPQ
AYAPAPVAPAPQAPPPQQPPQQQQWAPAPAAPPGAALLQSAGALAGPAPSAGQVAAAQ
AVLSRQAAPAAAPQAPQAPQFAPPPQAPQFAPAPQQPQPQQQQPAAPQFAPPPQAPAP
QPQFAPAPQQQQALAPQFAPAPQQPQQQPAFAPAPQAPAPQAPAPQFAPQQPLAPQFA
PVPQAPAYAPAPQQPSYAPASPPAAGSPQPRPPAGSPVRPPAPTTESVLSADSEEGDG
SYSFEFDASEASAASPVRVKGGRSGSIGAAASGSAGAPGSRSAAGRYGDVQQSLEALS
PAPRAPAAPAAAAGPPVLRRLPADPAAWAAAGAGDDTLILSVESLTLSADALRDPRLG

EGAHVMVSHMLLEAFTAARDQTTAAAPARRGAIGFEYAVGYSLLARDGAPSEAGAALQ
RQVAVEGECIVPLVLLATTNPQVDLDVTSSGEFDMIGMAEVDLSDWFVNPKREDWVSR
DVRLSPHESAAAGYPSGVAGVVRVTLRARGALAALRRACGEGR"

BASE COUNT 233 a 808 c 797 g 189 t 1 others
ORIGIN

Query Match 7.3%; Score 43.2; DB 15; Length 2028;
Best Local Similarity 43.8%; Pred. No. 19;
Matches 189; Conservative 0; Mismatches 243; Indels 0; Gaps 0;

```
Qy    137 cggtagtggcgacggcgaaaccggggtgtgtgtcctcgggcactgacgcggcgagtagca 196
Db    948 CGGGCGCACCGGCGAGCCGGCCGGCGGCCGCGGCTGCGGCGAGCCCGCCGCCGGCGGCGA 889

Qy    197 gtcataccaatacgacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgc 256
Db    888 CGCCGGCGCGTACGACGGCTGCTGCGGCGCCGGCGCGTAGGCGGGCGCCTGGGGCACCGG 829

Qy    257 cagccaaggcagccgcgcctcctgccgcggcgcgatcggctgagccgcacgtggggagca 316
Db    828 CGCGAACTGCGGCGCGAGCGGCTGCTGCGGCGCAAACTGCGGCGCCGGCGCCTGGGGGGC 769

Qy    317 agatcattgctaatctagtgaatcagctggggattaatgtcacccaaaggagcgtcgtca 376
Db    768 CGGCGCCTGCGGCGCCGGCGCGAACGCCGGCTGCTGTTGCGGCTGCTGCGGCGCCGGCGC 709

Qy    377 gcactggagcgccggccacgacgaggtctacggcggtgacgtccacgactaccgccccgc 436
Db    708 GAACTGGGGCGCAAGCGCCTGCTGCTGCTGCGGCGCCGGCGCGAACTGAGGCTGCGGCGC 649

Qy    437 agcgaacaagtccatacgggcacaatggccgacctgtgacggctggattggtggcagcta 496
Db    648 GGGCGCCTGCGGCGGCGGCGCGAACTGCGGCGCCGCCGGCTGCTGCTGCTGCGGCTGCGG 589

Qy    497 atagtggtgccagcgcggcctcgtctcccacagccgcggcgaaaccaacaggagaagaaa 556
Db    588 CTGCTGCGGGGCCGGCGCAAACTGCGGCGCCTGCGGCGGTGGCGCGAACTGCGGCGCCTG 529

Qy    557 aggcctccgcgg 568
Db    528 CGGCGCCTGCGG 517
```

RESULT 7
AC008532/c
LOCUS AC008532 158390 bp DNA HTG 18-JUL-2000
DEFINITION Homo sapiens chromosome 19 clone CTC-483I11, WORKING DRAFT SEQUENCE, 13 ordered pieces.
ACCESSION AC008532
VERSION AC008532.5 GI:9255996
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 158390)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 19
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 158390)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Jul 18, 2000 this sequence version replaced gi:7689764.
-------Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 370855, BC335879

```
            Center clone name: CIT-HSPC_483I11
            --------
            Summary Statistics
            Consensus quality: 145411 bases at least Q40
            Consensus quality: 154484 bases at least Q30
            Consensus quality: 156100 bases at least Q20
            Estimated insert size: 156970; agarose-fp estimation
            Estimated insert size: 157840; sum-of-contigs estimation
            Quality coverage:  5.76 in Q20 bases; agarose-fp estimation
            Quality coverage:  5.73 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1    10270: contig of 10270 bp in length
            *    10271    10370: gap of unknown length
            *    10371    17547: contig of 7177 bp in length
            *    17548    17647: gap of unknown length
            *    17648    62828: contig of 45181 bp in length
            *    62829    62928: gap of unknown length
            *    62929    74684: contig of 11756 bp in length
            *    74685    74784: gap of unknown length
            *    74785    87629: contig of 12845 bp in length
            *    87630    87729: gap of unknown length
            *    87730    88758: contig of 1029 bp in length
            *    88759    88858: gap of unknown length
            *    88859    96394: contig of 7536 bp in length
            *    96395    96494: gap of unknown length
            *    96495   126194: contig of 29700 bp in length
            *   126195   126294: gap of unknown length
            *   126295   136826: contig of 10532 bp in length
            *   136827   136926: gap of unknown length
            *   136927   138541: contig of 1615 bp in length
            *   138542   138641: gap of unknown length
            *   138642   142605: contig of 3964 bp in length
            *   142606   142705: gap of unknown length
            *   142706   156892: contig of 14187 bp in length
            *   156893   156992: gap of unknown length
            *   156993   158390: contig of 1398 bp in length.
FEATURES             Location/Qualifiers
     source          1. .158390
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="CTC-483I11"
                     /clone_lib="CalTech human BAC library C"
BASE COUNT    40327 a  39358 c  40000 g  37475 t   1230 others
ORIGIN


  Query Match             7.3%;  Score 43.2;  DB 60;  Length 158390;
  Best Local Similarity   45.0%;  Pred. No. 4.5;
  Matches  162;  Conservative    0;  Mismatches  198;  Indels    0;  Gaps    0;

Qy    124 ggccaaaaaacatcggtagtggcgacggcgaaaccggggtgtgtgtcctcgggcactgac 183
          |||   || ||  |||  | |||| ||| |      ||| |  |  | || ||      |
Db  87135 GGCATGAACACGCCGGAGGGGGCGGCGGTGGGGGGCGGGCGGCTTCCTTCAGGAGGGCGC 87076

Qy    184 gcggcgagtagcagtcataccaatacgacttctgccgctgctgcatcccctgcatcaccc 243
          ||     |       |   ||    || || |||| |  ||||   |  ||||  |  |
Db  87075 GCCCGCCGAGCGCCGCTCTCCGCGGCGGCTGCTGCTGTGGCTGTCGCTGCTGCTGCCGCT 87016

Qy    244 cctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggctgagccg 303
          |    | ||| | | ||    |  || | | || |||| ||  | ||  | || |   ||
Db  87015 CTAACTGCAGTGGCGGCTGCTGTGGCTGTGGCTGCTGCTGCTCCCCGGGCCGCAGGCGCG 86956

Qy    304 cacgtggggagcaagatcattgctaatctagtgaatcagctggggattaatgtcacccaa 363
            || |  | |   | |  | |     || |||   | || |  |      | | |||

Db  86955 TCCGCGTCGTGGCCGCTGCTGGGATCGCTGGTGCCGCCGCCGCCGCCGCCAGGCGCCCGC 86896

Qy    364 aggagcgtcgtcagcactggagcgccggccacgacgaggtctacggcggtgacgtccacg 423
            |   || |   || | |   | | |||| |   | ||    |||    |      |
Db  86895 TCGCATGTGGCTCGCGCCGCGTCTCAGGCCGCCCGGCGGATCCCGGGCCCGCTCCAAAGC 86836

Qy    424 actaccgccccgcagcgaacaagtccatacgggcacaatggccgacctgtgacggctgga 483
              |||||||||  ||  |    ||   | |||         || || |   | ||  ||
Db  86835 CGCCCCGCCCCGCCCCGCCCCGCCCCCGCCCGGCGGGTCCCACGCCCCGCCCCCGCGTGA 86776


RESULT   8
EIMMAX/c
LOCUS       EIMMAX        767 bp    mRNA            INV       19-MAY-1993
DEFINITION  E.tenella antigen LPMC61 mRNA, partial cds.
ACCESSION   M30933
VERSION     M30933.1  GI:158895
KEYWORDS    antigen.
SOURCE      E.tenella (strain Wisconsin), mRNA to cDNA.
  ORGANISM  Eimeria tenella
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
            Eimeria.
REFERENCE   1  (bases 1 to 767)
  AUTHORS   Ko,C., Smith,C.K.II. and McDonell,M.
  TITLE     Identification and characterization of a target antigen of a
            monoclonal antibody directed against Eimeria tenella merozoites
  JOURNAL   Mol. Biochem. Parasitol. 41, 53-64 (1990)
  MEDLINE   90348718
COMMENT     Draft entry and computer-readable sequence for [Unpublished (1990)
            Synergen, INC., 1885 33rd St., Boulder,CO 80301] kindly submitted
            by C.S.Ko 12-DEC-1989.
FEATURES             Location/Qualifiers
     source          1. .767
                     /organism="Eimeria tenella"
                     /db_xref="taxon:5802"
     CDS             <1. .>767
                     /note="antigen LPCM61"
                     /codon_start=1
                     /protein_id="AAA29079.1"
                     /db_xref="GI:158896"
                     /translation="LRLLLKLLLLLLGQQKHWPERQQQQQPQPWLDRQQQQQQHNQQL
                     QKQQWPEGQRQQLWPEQQQQQWPEQHQQAQQQQQWPQQQPQMQQEQWPQQQPQVQQQQ
                     QWPQQQHRRQHGQQQQCMNSQQQLQQCGQQQQQQLQQQWSEQQQQQQQQQQWPEQPEQQ
                     QQQQWPEQQQQQWSDQNQQQQAQQWQAQQQQQWPQQQQQPQQQQQQQQQQDLGPDGVG
                     IVVPYLGSSPADFEVQNLGVSVLPRIKLPLPKFDNDDG"
     repeat_region   52. .634
                     /note="region of degenerate repeats"
BASE COUNT      201 a    240 c    249 g     77 t
ORIGIN


  Query Match             7.2%;  Score 42.8;  DB 6;  Length 767;
  Best Local Similarity   56.3%;  Pred. No. 31;
  Matches   80;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy    145 gcgacggcgaaaccggggtgtgtgtcctcgggcactgacgcggcgagtagcagtcatacc 204
          ||   ||| |  | |    |||  |   || ||| |  || ||   | ||  |||  ||
Db    205 GCTCAGGCCACTGCTGCTGCTGTTGCTCTGGCCACAGCTGCTGCCGCTGGCCCTCAGGCC 146

Qy    205 aatacgacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaag 264
          | | |  |||||| |||| ||  |   ||||  |  |  |||||   |  | ||||| |
Db    145 ACTGCTGCTTCTGCAGCTGTTGGTTGTGCTGCTGCTGCTGCTGTTGTCGATCTAGCCACG 86

Qy    265 gcagccgcgcctcctgccgcgg 286
          || || |   || |||| || |
Db     85 GCTGCGGTTGCTGCTGCTGCTG 64


RESULT   9
AP001081/c
LOCUS       AP001081   140812 bp    DNA             PLN       12-AUG-2000
DEFINITION  Oryza sativa genomic DNA, chromosome 1, clone:P0693B08.
ACCESSION   AP001081 BA000010
```

```
VERSION     AP001081.1  GI:6815077
KEYWORDS    .
SOURCE      Oryza sativa (cultivar:Nipponbare) DNA, clone:P0693B08.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
            Oryza.
REFERENCE   1  (bases 1 to 140812)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
            clone:P0693B08
  JOURNAL   Published Only in DataBase (2000) In press
REFERENCE   2  (bases 1 to 140812)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JAN-2000) to the DDBJ/EMBL/GenBank databases. Takuji
            Sasaki, National Institute of Agrobiological Resources, Rice Genome
            Research Program; Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     The orientation of the sequence is from SP6 to T7 of the PAC clone.
            Genes were predicted from the integrated results of the
            following:GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as
            SplicePredictor (October1998 version). The genomic sequence was
            searched against the non-redundant database NRP(PIR,SWISSPROT,
            GENPEPT, PDB) from MAFF DNAbank and the cDNA sequence database at
            RGP. Protein similarities of the cording regions were searched
            against NRP with BLASTP2.0. ESTs represent the identified cDNA
            sequences using BLASTN2.0 with the corresponding DDBJ accession no.
            and RGP clone ID
            The sequence of this clone has an overlap with P0667A10 clone,
            DDBJ:AP001073 at the 5' end and an overlap with P0424A08 clone,
            DDBJ:AP000837 at the 3' end. This clone starts at the position
            106,268 of P0667A10 and ends at the position 3,914 of P0424A08.
            Detailed information on assemble quality together with annotation
            of this entry at http://www.dna.affrc.go.jp:82/genomicdata/GenomeFi
            nished.html.
FEATURES              Location/Qualifiers
     source           1. .140812
                      /organism="Oryza sativa"
                      /cultivar="Nipponbare"
                      /db_xref="taxon:4530"
                      /chromosome="1"
                      /clone="P0693B08"
     CDS              join(3816. .4050,4056. .4279,4294. .4770)
                      /note="hypothetical protein"
                      /codon_start=1
                      /protein_id="BAA90364.1"
                      /db_xref="GI:6815078"
                      /translation="MPHFSSLSLSLTGGSAAAGDHGRRAGAEAAASGGWAQGWRRPGM
                      EEEAAGGGRGAEAGDGQWRRRQRAAGGREAGDGREAIVGGSQGAEVAQAPAGGEEPRP
                      KEAACGERGSSNLLLAALMRARVQRAWRARREPTHGGGGGEAAATRGAAAPTLRSGGH
                      RAPTTRRRSPPPIARPRSPGRPPPPSPAAVVLEPPPPPAIGSVPARRRPHLTAATLLC
                      RLRLRADVHAVCAAVHTVLVDRPSPRHPRRPPAPTARPRRARCAPAFSTPASRGPRLL
                      DHRPPPPPPPPPVASLAPTRRPSPTSASRPLSPTHP"
     CDS              join(9496. .9723,10535. .11281)
                      /note="ESTs D24550(R2151),D24265(R1609),AU031848(R2151)
                      correspond to a region of the predicted gene.
                      Similar to cationic peroxidase isozyme 40K precursor
                      (D42065)"
                      /codon_start=1
                      /protein_id="BAA90365.1"
                      /db_xref="GI:6815079"
                      /translation="MDLAWWFAVAVVVCGLVGGGSAGLLETNPGLAYNFYQKSCPNVD
                      SIVRSVTWAQVAANPALPGRLLRLHFHDCFVQGCDASILLDNAGSEKTAGPNLSVGGY
                      EVIDAIKTQLEQACPGVVSCADIVALAARDAVSYQFKASLWQVETGRRDGPVSLASNT
                      GALPSPFAGFSTLLQSFANRGLNLTDLVALSGAHTIGKASCSSVTPRLYQGNTTSLDP
                      LLDSAYAKALMSSCPNPSPSSSTIDLDVATPLKFDSGYYANLQKKQGALASDAALTQN
                      AAAAQMVADLTNPIKFYAAFSMSMKKMGRIDVLTGSKGNIRKQCRSAS"
     CDS              join(13647. .13769,14188. .14391,14428. .14602,14993. .15468)
                      /note="hypothetical protein"
                      /codon_start=1
                      /protein_id="BAA90366.1"
                      /db_xref="GI:6815080"
                      /translation="MRVERVADLPVAMHQPTLLPVGSSSIATADTYEMNMQVGYRASR
                      AALLGTRPPGRGGRGGGGGSFDAGEGEGARLLRWRRVSCVSDKSGGGGESLLQASRIF
                      FLAILSPISASMYWSVSYTSRSGRAPPLGSAGPNPSFTTSLHTSIDLEYSRYQPRREL
                      VLVHSPAGPHMRRRDRRWLWAFAGAPRTGSTKTVRAQIIEQCTASPSCTLRKLALGAL
                      QLAEADHGAPRVRVFCIQPRGERLLHPQVDVRLHACRLYLGVPPPSVGVAAPCPGGLL
                      PSSAACLRREGKGEREMRKGENRGKREEADVGQADICNSCGSHADSAAT"
     CDS              join(16052. .16064,16146. .16319,16355. .16935)
                      /note="hypothetical protein"
                      /codon_start=1
                      /protein_id="BAA90367.1"
                      /db_xref="GI:6815081"
                      /translation="MGPTLGAAVEGERGDRWRSGDGGDRRGREVAGGGARVARKGKRG
                      RRNGELGQLAAAAGEETVGRGEERRGAWHGAAAWERSGVAPTSSLFERTLTNLLAGIP
                      ALLAGLLALLAGLPISFPSTPRASSPKRRSSTLEPPLCRSPVAATADHNDCLARTRKR
                      EGTYTTGRIWRSSEPAAAPIDTLRSTPLPRRCAVRCFSPRLPRAQVGRARTGLPAASP
                      SPSKPLCPTTRQPLPRSLAAHRLHPRIATPPAAGLSH"
     CDS              18583. .20061
                      /note="ESTs C71763(E0309),AU082059(E0309) correspond to a
                      region of the predicted gene.
                      Similar to Arabidopsis thaliana chromosome II BAC T20B5
                      genomic sequence; hypothetical protein. (AC002409)"
                      /codon_start=1
                      /protein_id="BAA90368.1"
                      /db_xref="GI:6815082"
                      /translation="MAIAVGHCRELSPPGGGVRGRHCRLRSVTQLDCPVFDDHDLDLD
                      RRERNGHAGLSPRRSLSSPCFSTVVPPAADRAEPARDDGEKMMPRVEIVAGGHARGVH
                      ELIAEAAGAIATGTRLVPAQGGIGGALLLEDGRSVDHVAVIKPLLDDASSPSHGGGGG
                      YASKAVLREVAAFLLDHDGFARVEPTALIKISRPAMPTTTASIQRFAAHECDAGELGP
                      SRFSVASVHRIGSLDVRLLNIDRHAGNILVKKSPESECASGGSTLTPLDLVPIDHGLC
                      LPEQLDDPYFEWLHWPQSSLPFSGAELEYVASLDPFRDAAMLRAELPSLTEAAIRILT
                      LCTIFLQRAAAAGLCLADIGDMMTREFSAMEEGLSALESLCKNAYDSSTTTAGSSPRK
                      HHHSDDSDDESTQFGMDDVPAGLPPHLFLLGGGGIAKSVSFSAAEQGAAGRGAARKRM
                      SFKALSGDEWAAFLDRFEQLLPAALDAKKRAGLKLTRLGTSF"
     LTR              complement(20793. .21346)
                      /note="3' LTR"
     CDS              join(21752. .21851,22552. .22676,22886. .23101)
                      /note="hypothetical protein"
                      /codon_start=1
                      /protein_id="BAA90369.1"
                      /db_xref="GI:6815083"
                      /translation="MDLDKSTSTSVSLEKLQEEGALPGRGTMEREPGGLNSETVGRRI
                      GQVMISGPTTASNVPVPLCENWAAEREAAINRPSHGRTARAHATRRHATAQARTAAAP
                      AASRSRSRRWSRAPPCGQPPAWGQTSTASVAAWGRVQGARDAAS"
     CDS              complement(join(23394. .23529,23558. .24018))
                      /note="hypothetical protein"
                      /codon_start=1
                      /protein_id="BAA90370.1"
                      /db_xref="GI:6815084"
                      /translation="MGWPARRRTRRRRRDLAQRLATASGGRSGGGGGGKEGDGARAV
                      ECTGAGDCKTKKREEGTGVLYIGSRGLIVAGVAAISPAEWREVGRRERRDSKIESRPS
                      RARTRAEEREWARGRRAQAATWATWPERRGSERGGGCGMQRRRLEVGDGPDREEGGGR
                      LGWAVAFGWPSAEGERKRMGRRPIRKRKRREKKIKRRK"
     CDS              complement(join(24161. .24471,24526. .24637))
                      /note="hypothetical protein"
                      /codon_start=1
                      /protein_id="BAA90371.1"
                      /db_xref="GI:6815085"
                      /translation="MAVTRVAVHDEPGPRTGEGAVHHVALTWRDVAHTWQPHRTAAGS
                      ARLARTGSSSRPGSNGVRRRYRGAAAADGGEWRRRDERRRRDATTRRRRYRRKREGEE
                      GGGVLTEEVRPAWRETTVNGGETRQREVDKRRTGSGTT"
     CDS              join(25452. .25542,26685. .26854,26925. .26983,27057. .27489,
                      27514. .27825)
                      /note="EST AU082058(C12976) corresponds to a region of the
                      predicted gene.
                      hypothetical protein"
                      /codon_start=1
                      /protein_id="BAA90372.1"
                      /db_xref="GI:6815086"
                      /translation="MERGGPYDGLCRFRIHLGTRGDCPLPAGDGVGPLVDHQAAASLK
                      EDVAKEASDAAAAVATTSGVNVPKKGRKFSSVLGNRRKTPTPSASDASPPPPRRQRLV
                      TLGEKAARAKAAQDRSGGNSSASPVVASTDVVVMTRSREATPNGPVGDLVPTQGSPAD
                      ILTWEELQVEMGRILQAGARGIRHEIAEARAAAASSANERADKLARDLAEVREDLQKM
```

```
                RELVAGNERQRQGLKNRMSELENNLSEIRGSLRLARECGVTTTIPANPDEFLLTTSLA
                ELATAMEVIPSKHAARIGEETSSGIYTGTCNVLACVKLAHPELDLQKILYQGAASNAR
                KDVMEEIGDMGESILPLFEE"
     CDS        complement(join(29314. .30354,33082. .35175))
                /note="Similar to GAG-POL precursor (AB030283)
                internal stop
                codon:complement(29869. .29871,30130. .30132,30295. .30297,
                34924. .34926,34960. .34962)
                probably inactive because stop codons are included in CDS"
                /codon_start=1
                /pseudo
     LTR        complement(30355. .30714)
                /note="3' LTR"
     CDS        complement(30656. .31903)
                /note="Similar to Zea mays 22-kDa alpha zein gene cluster;
                hypothetical protein. (AF031569)"
                /codon_start=1
                /protein_id="BAA90373.1"
                /db_xref="GI:6815087"
                /translation="MAEPAKAHDLSPSMSGDDGEPNPRRRARTPPPPPRQSPRREKAL
                ERAEGSATSTSTGGGEGRRDGERRLLVYGDGSTPQGALQAAGALLRHPPVAHDPESPA
                QRWLDDVAKLVMTARQRLDAGGRSSATKASGAATTGSASSRRRARRAAAAVRHSATTP
                SSTPSTREDLRGGPDARTSIERRRNGRRATHATEGASSSRVSPQHGRRNQPSVPPVGG
                VGCRAFVASLRNVRWPPRFRPTIAEKYDGSVNPAEFLQVYTTGIEAAGGDDRVMANFF
                PMALKGQARGWLMNLPPASVHSWEDLCQQFTMKFQGTYPRPGEEADLHAVQRRDDESL
                RSYIQRFCQVRNTIPCIPTHAVIYAFRGGVRHNRMLEKIASKEPQTTAELFQLADRVA
                RMKRWGKSPQKRGKWCMWYPLKI"
     LTR        complement(31991. .32353)
                /note="5' LTR"
     LTR        complement(38386. .38937)
                /note="5' LTR"
     CDS        join(40089. .40184,40259. .40396,40823. .41078,41886. .41957,
                42532. .42671)
                /note="ESTs C91695(E31262),D40183(S1971) correspond to a
                region of the predicted gene.
                hypothetical protein"
                /codon_start=1
                /protein_id="BAA90374.1"
                /db_xref="GI:6815088"
                /translation="MAASAPPPAEPPPPPAAAAAAATNPPASTAPTEDGEVEEGATPP
                RTSALGTANWGTATLVGVFAGLLYGGSKEASASVSKDAEVMLKMGSTTDKREQYRLMR
                DAMEKRFIRVAKGSLVGGVRLGMFTATFFGIQNLLIEKRGVHDVFNIAGAGSATAAAF
                GLICSPMWRARNVLVGSALGAGVCFPLGWIQLKLAEKANLEMASSKPTDLVEEQGNQS
                GVGAAIERLGRSLKK"
     CDS        complement(join(43206. .43424,44141. .44605,44699. .45774,
                45861. .46014,46110. .46206,46565. .46701))
                /note="Similar to glutamyl-tRNA synthetase (AF067773)"
                /codon_start=1
                /protein_id="BAA90375.1"
                /db_xref="GI:6815089"
                /translation="MEPKLAFPQDSPPLAILCAAKVAGVSLTLDPKLASGSAPTLHLG

  Query Match             7.2%;  Score 42.6;  DB 13;  Length 140812;
  Best Local Similarity   48.2%;  Pred. No. 6.2;
  Matches  120;  Conservative    0;  Mismatches  129;  Indels    0;  Gaps    0;

Qy    161 ggtgtgtgtcctcgggcactgacgcggcgagtagcagtcataccaatacgacttctgccg 220
          | |  | | || |||   | |  ||||      ||        ||   ||||  | || |
Db  72480 GTTCGGCGGCCCCGGCGGCCGCGGCGGACCCGCGCTCGGCCCTCACGGCGACGACGGCGG 72421

Qy    221 ctgctgcatcccctgcatcacccccctgtttcagcgccagccaaggcagccgcgcctcctg 280
          |  | ||  | || ||  |   | | |  || | | |  | |||| || ||| |  | |
Db  72420 CACCGGCGGCACCGGCGGCTTTCTCGGCGTCGGAGGCGACGGAGGCCGCGGCGGCGACGG 72361

Qy    281 ccgcggcgcgatcggctgagccgcacgtggggagcaagatcattgctaatctagtgaatc 340
          | ||   |    | || ||| |    | || ||   || |||  ||  | || |  ||
Db  72360 CGGCCCAGGAGGCAGCGGAGGCCGCGGCGGCGACGGAGGTCAAGGCGGACCTGGCCAACG 72301

Qy    341 agctggggattaatgtcacccaaaggagcgtcgtcagcactggagcgccggccacgacga 400
           |  |||||    | | |    ||   |  | |    | ||  || |||| |||  |
Db  72300 GGGCGGGGACGGCGGGCCCGGCGGGGCAGGCGGGCCCGGCGGGCACGGCGGCGACGGTGG 72241

Qy    401 ggtctacgg 409
```

```
          || |  |||
Db  72240 GGACGGCGG 72232


RESULT  10
HSE4
LOCUS       HSE4          5880 bp    DNA             VRL       03-FEB-1999
DEFINITION  Equine herpesvirus 4 (EHV4) genome, thymidine kinase (TK) and
            glycoprotein H (gH) genes.
ACCESSION   D14486 D00683 D00684
VERSION     D14486.1  GI:221820
KEYWORDS    UL21 gene; UL24 gene; gH gene; glycoprotein; thymidine kinase.
SOURCE      Equine herpesvitus 4 ( strain 1942).
  ORGANISM  Equine herpesvirus 4
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Alphaherpesvirinae; Varicellovirus.
REFERENCE   1  (sites)
  AUTHORS   Nicolson,L., Cullinane,A.A. and Onions,D.E.
  TITLE     The nucleotide sequence of an equine herpesvirus 4 gene homologue
            of the herpes simplex virus 1 glycoprotein H gene
  JOURNAL   J. Gen. Virol. 71 (Pt 8), 1793-1800 (1990)
  MEDLINE   90362066
REFERENCE   2  (sites)
  AUTHORS   Nicolson,L., Cullinane,A.A. and Onions,D.E.
  TITLE     The nucleotide sequence of the equine herpesvirus 4 thymidine
            kinase gene
  JOURNAL   J. Gen. Virol. 71 (Pt 8), 1801-1805 (1990)
  MEDLINE   90362067
REFERENCE   3  (bases 1 to 5880)
  AUTHORS   Nicolson,L.
  JOURNAL   Unpublished (1993)
COMMENT     Submitted (10-SEP-1990) to DDBJ by:
            Lesley Nicolson
            Dept. Veterinary Pathology. University of Glasgow
            Vet School
            Bearsden Roda, Glasgow G61 1QH. Scotland
            UK.
            Phone: 041-339-8855
            Fax:   041-330-5733.
FEATURES             Location/Qualifiers
     source          1. .5880
                     /organism="Equine herpesvirus 4"
                     /db_xref="taxon:10331"
     polyA_signal    complement(99. .104)
                     /note="put. polyadenylation signal (UL24); putative"
     polyA_signal    complement(115. .120)
                     /note="put. polyadenylation signal (UL24); putative"
     CDS             complement(116. .934)
                     /note="putative;  ORF1 (UL24 homologue)"
                     /codon_start=1
                     /protein_id="BAA03377.1"
                     /db_xref="GI:221821"
                     /translation="MKRKQRLTARSRLRAGIRCHNRFYNAMVQDLASAKKNGVYGARL
                     APLFSELVPAETLKTAMGVSLAFEVNLGQRRPDCVCTVQFGHGSDAKGVCILIELKTC
                     RFSKNMNTASKNLQRKGGMRQLHDSCRLLAKTLPPGSGEILLAPVLVFVAQRGMRVLR
                     VTRLSPQVVYSNAAVLSCTISRLAEYSPPISERSTRRRCVTRRTNSKAFRAKTTTGSI
                     QPITQAKPAATAAVASLFSATAQANTTNAAVGYQPATISLANPLAWVASLFAPK"
     TATA_signal     861. .865
                     /note="put. TATA box (TK); putative"
     misc_feature    887
                     /note="pot. transcription initiation site; putative"
     CDS             949. .2007
                     /partial
                     /note="ORF2;  (TK)"
                     /codon_start=1
                     /protein_id="BAA03378.1"
                     /db_xref="GI:221822"
                     /translation="MAACVPPGEAPRSASGTPTRRQVTIVRIYLDGVYGIGKSTTGRV
                     MASAASGGSPTLYFPEPMAYWRTLFETDVISGIYDTQNRKQQGNLAVDDAALITAHYQ
                     SRFTTPYLILHDHTCTLFGGNSLQRGTQPDLTLVFDRHPVASTVCFPAARYLLGDMSM
                     CALMAMVATLPREPQGGNIVVTTLNVEEHIRRLRTRARIGEQIDITLIATLRNVYFML
                     VNTCHFLRSGRVWRDGWGELPTSCGAYKHRATQMDAFQERVSPELGDTLFALFKTQEL
                     LDDRGVILEVHAWALDALMLKLRNLNVFSADLSGTPRQCAAVVESLLPLMSSTLSDFD
```

```
                SASALERAARTFNAEMGV"
     polyA_signal    2049. .2054
                     /note="put. polyadenylation signal (TK); putative"
     TATA_signal     2127. .2133
                     /note="putative"
     TATA_signal     2197. .2202
                     /note="putative"
     CDS             2225. .4792
                     /partial
                     /note="ORF3;  (gH)"
                     /codon_start=1
                     /protein_id="BAA03379.1"
                     /db_xref="GI:221823"
                     /translation="MSQPYLKIAILVAATIVSAIPVWTTPVSTSPPQQTKLHYVGNGT
                     WVHNNTFNVTRYDRITMEPVYNNNLSSTTFFVAISERNFRTVNTPLGASVFWILKSAL
                     NPPKHQPCIANVPEPGDPRGPCVNSTVSLFFNDNLEPFLMTKNLLEFEVLPDNYITGW
                     TFERSKTVATKGNPVGVVLSPPRTSPDVNNTIRDDGTPKQHLSIIDEHTTFVLDLQNF
                     TKTLTYISPFAAVWPITAFHAGITVMGCDTTQAIAYLGNGFMGLQISSVNNPPLEMIV
                     APNDVRARIVNRLPPRRRLEPPGPYAGPIYKVYVLSDGNFYLGHGMSKISREVAAYPE
                     ESLDYRYHLSLANLDTLAMLAELSSGKSKDVSYYLYRIIARLAVATFSLAEVIRLSDY
                     MLLQEAIDVDINLRLIVPLVMKYAAGGTADSSYTSSDVAMDQFEVAQAQIEKIVADIN
                     IENELRKPMYEHRSLLKSVYAYSRKPLPNAVSFANRLITAMYKEAIKDRITWNSTMRE
                     VLFFAVGAAAGSHVILTDGPDLGLHAHKDSSMFLSLNRNILLLCTAMCTASHAVSAGV
                     KLEEVMAGLIAGGVQFSLLEVFSPCMASARFDLAEEEHVLDLLSVIPPRLYTDLNTGL
                     EDDGTTIHSYGRSANGILNSRIAYNFDAVRVFTPELASCSTKLPKVLVVLPLASNRSY
                     VITRTAPNIGLTYSLDGVNIAKPIVISYITYGNCQVSRATIRSVYLDHPGHTQSCVYC
                     GSVFMRYMASGAIMDLIYIDDKDVELQLVAGENSTIPAFNPKLYTPSMNALLMFPNGT
                     VTLMSAFASYSAFKIPSTYLWASIGGLLLAILILYVIVKMLCGGVINNDYSLLLNSE"
     polyA_signal    4777. .4782
                     /note="put. polyadenylation signal(gH); putative"
     repeat_region   4893. .5315
                     /note="direct repeat region"
     rep_origin      5395. .5403
                     /note="put. replication origin; putative"
     rep_origin      5440. .5448
                     /note="put. replication origin; putative"
     CDS             complement(5755. .>5880)
                     /note="putative;  ORF4 (UL21 homologue)"
                     /codon_start=1
                     /protein_id="BAA03380.1"
                     /db_xref="GI:221824"
                     /translation="NVSLITAFDNLDLSRKGAYYLYYLLSERLKRGGVPVHVNRA"
BASE COUNT     1459 a   1377 c   1499 g   1545 t
ORIGIN
```

```
  Query Match             7.0%;  Score 41.6;  DB 59;  Length 5880;
  Best Local Similarity   44.1%;  Pred. No. 28;
  Matches  173;  Conservative    0;  Mismatches  219;  Indels    0;  Gaps    0;

Qy    145 gcgacggcgaaaccggggtgtgtgtcctcgggcactgacgcggcgagtagcagtcatacc 204
          ||| ||||||   | |     | | |   ||   |||  ||||||||           |
Db   4893 GCGGCGGCGAGGGCTGCTGCGGCGGCGAGGGCTGCTGCGGCGGCGAGGGCTGCGGCGGCG 4952

Qy    205 aatacgacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaag 264
             | | || | || || || ||     ||||  |  |  | |     ||  | ||   |
Db   4953 GCGAGGGCTGCGGCGGCGGCGGCGAGGGCTGCGGCGGCGGCGGCGAGGGCTGCGGCGGCG 5012

Qy    265 gcagccgcgcctcctgccgcggcgcgatcggctgagccgcacgtggggagcaagatcatt 324
          || ||   | || | || ||||||     |||||   ||   | || |||
Db   5013 GCGGCGAGGGCTGCGGCGGCGGCGGCGAGGGCTGCTGCGGCGGCGGCGAGGGCTGCTGCG 5072

Qy    325 gctaatctagtgaatcagctggggattaatgtcacccaaaggagcgtcgtcagcactgga 384
          ||          |  |    || |           |       |   |  |  ||  |||
Db   5073 GCGGCGGCGAGGGCTGCTGCGGCGGCGGCGAGGGCTGCTGCGGCGGCGGCGAGGGCTGCT 5132

Qy    385 gcgccggccacgacgaggtctacggcggtgacgtccacgactaccgccccgcagcgaaca 444
          ||| ||||  ||| |    || ||||||  | ||    |  || | ||  ||  | |  |
Db   5133 GCGGCGGCGGCGAGGGCTGCTGCGGCGGCGGCGAGGGCTGCTGCGGCGGCGGCGAGGGCT 5192

Qy    445 agtccatacgggcacaatggccgacctgtgacggctggattggtggcagctaatagtggt 504
            | |    | |   |  |       ||| | |||| |    ||  || |||      |
Db   5193 GCTGCGGCGGCGGCGAGGGCTGCTGCTGCGGCGGCGGCGAGGGCTGCTGCTGCGGCGGCG 5252

Qy    505 gccagcgcggcctcgtctcccacagccgcggc 536
            || || || ||  || | |  |  |    || ||
Db   5253 GCGAGGGCTGCTGCGGCGGCGGCGAGGGCTGC 5284
```

```
RESULT  11
RNU87960/c
LOCUS       RNU87960     2479 bp    DNA             ROD       04-MAR-1997
DEFINITION  Rattus norvegicus leukocyte common antigen receptor (LAR) gene,
            trans-spliced alternative untranslated exon.
ACCESSION   U87960
VERSION     U87960.1  GI:1857717
KEYWORDS    .
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 2479)
  AUTHORS   Zhang,C., Zhang,J.S., Martignetti,J.A., Massa,S.M. and Longo,F.M.
  TITLE     Evidence for a mammalian trans-spliceon
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2479)
  AUTHORS   Zhang,C., Zhang,J.S., Martignetti,J.A., Massa,S.M. and Longo,F.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JAN-1997) Neurology, Veterans Administration Medical
            Center at UC San Francisco, Neurology (127) 4150 Clement St., San
            Francisco, CA 94121, USA
FEATURES             Location/Qualifiers
     source          1. .2479
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
     intron          <1. .1649
                     /gene="LAR"
     gene            1. .2479
                     /gene="LAR"
     repeat_region   1647. .2456
                     /rpt_type=tandem
                     /rpt_unit=CAG
     exon            1650. .2456
                     /gene="LAR"
                     /note="trans-spliced alternative untranslated exon"
                     /product="leukocyte common antigen receptor"
     intron          2457. .2479
                     /gene="LAR"
BASE COUNT      832 a    504 c    584 g    559 t
ORIGIN
```

```
  Query Match             6.9%;  Score 41.2;  DB 95;  Length 2479;
  Best Local Similarity   57.9%;  Pred. No. 46;
  Matches   73;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy    173 cgggcactgacgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatccc 232
          | |||||||  || |   |       | |   | |  || |||| ||||| ||  |||
Db   2186 CTGGCACTGCTGCTGTAGCTGCTGCAGCTGCTGCTGCTGCTGCTGCAGCTGCGGCTGCCC 2127

Qy    233 ctgcatcacccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgat 292
          |||||  |  || ||| | |  || | |||   || || ||  || |||||| |||
Db   2126 CTGCAGCTGCCGCTGGTGCCTCGGCTGCCCCTGCTGCTGCCGCTGCTGCCGAGGCTGCCC 2067

Qy    293 cggctg 298
          | ||||
Db   2066 CTGCTG 2061
```

```
RESULT  12
E34519/c
LOCUS       E34519       477 bp    DNA             PAT       07-FEB-2001
DEFINITION  SCA7 gene and utilization thereof.
ACCESSION   E34519
VERSION     E34519.1  GI:13018887
```

```
KEYWORDS    JP 1999206393-A/1.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 477)
  AUTHORS   Laura,B.W.R.R. and D,K.
  TITLE     SCA7 gene and utilization thereof
  JOURNAL   Patent: JP 1999206393-A 1 03-AUG-1999;
            THE REGENTS OF THE UNIVERSITY OF MINNESOTA
COMMENT     OS   Homo sapiens (human)
            PN   JP 1999206393-A/1
            PD   03-AUG-1999
            PF   19-AUG-1998 JP 1998294732
            PR   19-AUG-1997 US    60/056170
            PI   LAURA B W RANAMU,MICHAEL D KUBU
            PC   C12N15/09,C07K14/47,C07K16/18,C12Q1/68,G01N33/53, PC
            G01N33/566//C12P21/02,
            PC   C12N15/00
            CC
            FH   Key             Location/Qualifiers
            FT   source          1. .477
            FT                   /organism='Homo sapiens (human)'.
FEATURES              Location/Qualifiers
     source           1. .477
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
BASE COUNT       99 a    164 c    172 g     42 t
ORIGIN
```

```
  Query Match             6.8%;  Score 40.4;  DB 10;  Length 477;
  Best Local Similarity   53.9%;  Pred. No. 1.2e+02;
  Matches   83;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy     145 gcgacggcgaaaccggggtgtgtgtcctcgggcactgacgcggcgagtagcagtcatacc 204
           | | |||||    ||||   ||   || | ||  |||  | ||||        |  | |
Db     334 GTGGCGGCGGTGGCGGGTGCTGCTGCCGCTGGGGCTGCGGAGGCGGCGGCTGCTGCTGCT 275

Qy     205 aatacgacttctgccgctgctgcatcccctgcatcacccctgtttcagcgccagccaag 264
           | |  || |||| ||||||||  |  ||||  |  |  ||| | | ||  | ||
Db     274 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 215

Qy     265 gcagccgcgcctcctgccgcggcgcgatcggctg 298
           || ||  ||  || |||| ||  ||     | ||||
Db     214 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 181
```

```
RESULT  13
AF020275/c
LOCUS       AF020275      477 bp    DNA             PRI       15-OCT-1997
DEFINITION  Homo sapiens expanded SCA7 CAG repeat.
ACCESSION   AF020275
VERSION     AF020275.1  GI:2501955
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 477)
  AUTHORS   Koob,M.D., Benzow,K.A., Bird,T.D., Moseley,M.L. and Ranum,L.P.W.
  TITLE     Rapid cloning of expanded trinucleotide repeat sequences
  JOURNAL   Nature Genet. (1997) In press
REFERENCE   2  (bases 1 to 477)
  AUTHORS   Koob,M.D., Benzow,K.A., Moseley,M.L. and Ranum,L.P.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-AUG-1997) Neurology, University of MN, Box 295 UMHC,
            416 Delaware St. SE, Minneapolis, MN 55455, USA
FEATURES              Location/Qualifiers
     source           1. .477
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="3"
                      /map="3p"
                      /note="isolated from a spinocerebellar ataxia 7 (SCA7)
                      patient"
     misc_feature     115. .477
                      /note="similar to NotI jumping clone sequence with GenBank
                      Accession Number X95831"
     repeat_region    127. .285
                      /note="region of cag expansion"
                      /rpt_unit=cag
BASE COUNT       99 a    164 c    172 g     42 t
ORIGIN
```

```
  Query Match             6.8%;  Score 40.4;  DB 88;  Length 477;
  Best Local Similarity   53.9%;  Pred. No. 1.2e+02;
  Matches   83;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy     145 gcgacggcgaaaccggggtgtgtgtcctcgggcactgacgcggcgagtagcagtcatacc 204
           | | |||||    ||||   ||   || | ||  |||  | ||||        |  | |
Db     334 GTGGCGGCGGTGGCGGGTGCTGCTGCCGCTGGGGCTGCGGAGGCGGCGGCTGCTGCTGCT 275

Qy     205 aatacgacttctgccgctgctgcatcccctgcatcacccctgtttcagcgccagccaag 264
           | |  || |||| ||||||||  |  ||||  |  |  ||| | | ||  | ||
Db     274 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 215

Qy     265 gcagccgcgcctcctgccgcggcgcgatcggctg 298
           || ||  ||  || |||| ||  ||     | ||||
Db     214 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 181
```

```
RESULT  14
HSM801297/c
LOCUS       HSM801297    3932 bp    mRNA            PRI       18-FEB-2000
DEFINITION  Homo sapiens mRNA; cDNA DKFZp434H177 (from clone DKFZp434H177);
            partial cds.
ACCESSION   AL133030
VERSION     AL133030.1  GI:6453431
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3932)
  AUTHORS   Koehrer,K., Beyer,A., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-1999) MIPS, Am Klopferspitz 18a, D-82152
            Martinsried, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by BMFZ (Biomedical Research Center at the Charite,
            Berlin/Germany) within the cDNA sequencing consortium of the German
            Genome Project.
            This clone (DKFZp434H177) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
            information about the clone and the sequencing project is available
            at http://www.mips.biochem.mpg.de/proj/cDNA/.
FEATURES              Location/Qualifiers
     source           1. .3932
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="DKFZp434H177"
                      /clone_lib="434 (synonym: htes3). Vector pSport1; host
                      DH10B; sites NotI + SalI"
                      /dev_stage="adult"
                      /tissue_type="testis"
     gene             1. .3505
                      /gene="DKFZp434H177"
     CDS              <1. .3505
                      /gene="DKFZp434H177"
                      /note="similarity to peripheral benzodiazepine receptor
                      interactingprotein"
                      /codon_start=2
```

```
                     /product="hypothetical protein"
                     /protein_id="CAB61362.1"
                     /db_xref="GI:6453432"
                     /translation="SEQASAPLAKDKQIEELRQECHLLQARVASGPCSDLHTGRGGPC
                     TQWLNVRDLDRLQRESQREVLRLQRQLMLQQGNGGAWPEAGGQSATCEEVRRQMLALE
                     RELDQRRRECQELGAQAAPARRRGEEAETQLQAALLKNAWLAEENGRLQAKTDWVRKV
                     EAENSEVRGHLGRACQERDASGLIAEQLLQQAARGQDRQQQLQRDPQKALCDLHPSWK
                     EIQALQCRPGHPPEQPWETSQMPESQVKGSRRPKFHARPEDYAVSQPNRDIQEKREAS
                     LEESPVALGESASVPQVSETVPASQPLSKKTSSQSNSSSEGSMWATVPSSPTLDRDTA
                     SEVDDLEPDSVSLALEMGGSAAPAAPKLKIFMAQYNYNPFEGPNDHPEGELPLTAGDY
                     IYIFGDMDEDGFYEGELEDGRRGLVPSNFVEQIPDSYIPGCLPAKSPDLGPSQLPAGQ
                     DEALEEDSLLSGKAQGVVDRGLCQMVRVGSKTEVATEILDTKTEACQLGLLQSMGKQG
                     LSRPLLGTKGVLRMAPMQLHLQNVTATSANITWVYSSHRHPHVVYLDDREHALTPAGV
                     SCYTFQGLCPGTHYRARVEVRLPRDLLQVYWGTMSSTVTFDTLLAGPPYPPLDVLVER
                     HASPGVLVVSWLPVTIDSAGSSNGVQVTGYAVYADGLKVCEVADATAGSTLLEFSQLQ
                     VPLTWQKVSVRTMSLCGESLDSVPAQIPEDFFMCHRWPETPPFSYTCGDPSTYRVTFP
                     VCPQKLSLAPPSAKASPHNPGSCGEPQAKFLEAFFEEPPRRQSPVSNLGSEGECPSSG
                     AGSQAQELAEAWEGCRKDLLFQKSPQNHRPPSVSDQPGEKENCYQHMGTSKSPAPGFI
                     HLRTECGPRKEPCQEKAALERVLRQKQDAQGFTPPQLGASQQYASDFHNVLKEEQEAL
                     CLDLWGTERREERREPEPHSRQGQALGVKRGCQLHEPSSALCPAPSAKVIKMPRGGPQ
                     QLGTGANTPARVFVALSDYNPLVMSANLKAAEEELVFQKRQLLRVWGSQDTHDFYLSE
                     CNRQVGNIPGRLVAEMEVGTEQTDRRWRSPAQGHLPSVAHLEDFQGLTIPQGSSLVLQ
                     GNSKRLPLWTPKIMIAALDYDPGDGQMGGQGKGRLALRAGDVVMVYGPMDDQGFYYGE
                     LGGHRGLVPAHLLDHMSLHGH"
     polyA_site      3884
BASE COUNT      895 a   1143 c   1203 g    691 t
ORIGIN
```

```
  Query Match             6.8%;  Score 40.4;  DB 93;  Length 3932;
  Best Local Similarity   52.9%;  Pred. No. 58;
  Matches  109;  Conservative    0;  Mismatches   96;  Indels    1;  Gaps    1;

Qy     141 agtggcgacggcgaaaccggggtgtgtgtcctcgggcactgacgcggcgagtagcagtca 200
           ||  ||| || |   | | | || |  | |  |  ||   | ||| |||    |    |
Db     414 AGCAGCGCCGCCTGCAGCTGTGTCTCGGCCTCCTCGCCACGTCGCCGCGCCGGGGCCGCC 355

Qy     201 taccaatacgacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgccagc 260
           | |     |  || ||| | ||   ||  || |||   || | |  | | ||||||||||
Db     354 TGCGCGCCCAGCTCCTGGCACTCGCGCCGCCGCTGGTCCAGCTCGCGCTCCAGCGCCAGC 295

Qy     261 caaggcagccgcgcctcctgccgcggcgcgatc-ggctgagccgcacgtggggagcaaga 319
           || | ||| |||||  |  |  ||| || ||| |  |  | || |   |||   |
Db     294 ATCTGCCGTCGCACCTCCTCGCAGGTTGCGCTCTGGCCGCCCGCCTCGGGCCAAGCGCCA 235

Qy     320 tcattgctaatctagtgaatcagctg 345
           | ||||    ||   | |||| |||
Db     234 CCGTTGCCCTGCTGAAGCATCAACTG 209
```

```
RESULT  15
RNLAR2/c
LOCUS       RNLAR2       3941 bp    mRNA            ROD       27-FEB-1995
DEFINITION  R.norvegicus mRNA for leucocyte common antigen-related protein
            (3941 bp).
ACCESSION   X83546
VERSION     X83546.1  GI:673488
KEYWORDS    cell adhesion molecule; leucocyte common antigen-related protein;
            nerve growth factor; trinucleotide repeat; tyrosine phosphatase.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 3941)
  AUTHORS   Zhang,J.S. and Longo,F.M.
  TITLE     LAR tyrosine phosphatase receptor: alternative splicing is
            preferential to nervous system, coordinated with cell growth and
            generates novel isoforms containing extensive CAG repeats
  JOURNAL   J. Cell Biol. (1995) In press
REFERENCE   2  (bases 1 to 3941)
  AUTHORS   Zhang,J.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-DEC-1994) J.S. Zhang, Univ. of California San
            Fransisco, Neurology, 4150 Clement Street, Building 1 Room 219H,
            San Francisco CA 94121, USA
FEATURES             Location/Qualifiers
     source          1. .3941
                     /organism="Rattus norvegicus"
                     /strain="Sprague-dawley"
                     /db_xref="taxon:10116"
                     /clone="pRLAR4.0"
                     /dev_stage="adult"
                     /tissue_type="brain"
                     /clone_lib="cDNA library (custom made by Strategene)"
     CDS             <1. .1863
                     /codon_start=1
                     /product="leucocyte common antigen-related protein"
                     /protein_id="CAA58537.1"
                     /db_xref="GI:693993"
                     /db_xref="SPTREMBL:Q63296"
                     /translation="SIGLKDSLLAHSSDPVEMRRLNYQTPGSSAPSCPNISSMRDHPP
                     IPITDLADNIERLKANDGLKFSQEYESIDPGQQFTWENSNSEVNKPKNRYANVIAYDH
                     SRVLLTSIDGVPGSDYINANYIDGYRKQNAYIATQGPLPETMGDFWRMVWEQRTATVV
                     MMTRLEEKSRVKCDQYWPARGTETYGLIQVTLVDTVELATYTMRTFALHKSGSSEKRE
                     LRQFQFMAWPDHGVPEYPTPILAFLRRVKACNPLDAGPMVVHCSAGVGRTGCFIVIDA
                     MLERMKHEKTVDIYGHVTCMRSQRNYMVQTEDQYVFIHEALLEAAMCGHTEVLARNLY
                     AHIQKLGQVPPGESVTAMELEFKLLANSKAHTSRFISANLPCNKFKNRLVNIMPYELT
                     RVCLQPIRGVEGSDYINASFLDGYRQQKAYIATQGPLAESTEDFWRMLWEHNSTIIVM
                     LTKLREMGREKCHQYWPAERSARYQYFVVDPMAEYNMPQYILREFKVTDARDGQSRTI
                     RQFQFTDWPEQGVPKTGEGFIDFIGQVHKTKEQFGQDGPITVHCSAGVGRTGVFITLS
                     IVLERMRYEGVVDMFQTVKTLRTQRPAMVQTEDQYQLCYRAALEYLGSFDHYAT"
     misc_feature    80. .112
                     /note="insertion sequence LASEa"
     3'UTR           2104. .3941
                     /note="alternate spliced 3'UTR with CAG repeats"
BASE COUNT     1001 a   1069 c   1061 g    810 t
ORIGIN
```

```
  Query Match             6.8%;  Score 40.4;  DB 95;  Length 3941;
  Best Local Similarity   56.9%;  Pred. No. 58;
  Matches   74;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy     173 cgggcactgacgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatccc 232
           | | | |||  |  |||  |  |  |  | |   | |  || |||| |||||||| |
Db    2718 CTGCCGCTGCAGTCGCGGTTGCCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCAGCTG 2659

Qy     233 ctgcatcacccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgat 292
           | ||  | ||  ||| | |  |    |||  ||| ||| |  || |||||| ||
Db    2658 CGGCTGCCCCCTGCTGCTGCCTCTGATGCCTCGGCTGCCCCTGCTGCTGCCGCTGCTGCCG 2599

Qy     293 cggctgagcc 302
           ||||||  ||
Db    2598 CGGCTGCCCC 2589
```

```
Search completed: September 21, 2001, 00:27:43
Job time: 29118 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: September 20, 2001, 23:31:44 ; Search time 330.77 Seconds
(without alignments)
1127.592 Million cell updates/sec

Title: US-09-138-735-1_COPY_1232_1825
Perfect score: 594
Sequence: 1 caggtacagcgtaacggctt.....aaacgagctccgtggcgata 594

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0601:*
1: /cgn1_9/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /cgn1_9/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /cgn1_9/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /cgn1_9/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /cgn1_9/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /cgn1_9/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /cgn1_9/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /cgn1_9/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /cgn1_9/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /cgn1_9/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /cgn1_9/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /cgn1_9/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /cgn1_9/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /cgn1_9/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /cgn1_9/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /cgn1_9/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /cgn1_9/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /cgn1_9/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	594	100.0	3402	20	AAX84092	T. cruzi PTc40 cod
	2	590.8	99.5	3402	17	AAT27310	Trypanosoma cruzi
c	3	40.4	6.8	477	21	AAZ44307	Human SCA7 genomic
c	4	39.8	6.7	567	21	AAA29550	HIV codon altered
	5	39.2	6.6	2790	19	AAV18480	BOP1 cDNA. Mus sp
c	6	38.2	6.4	2943	17	AAT16480	SA8 virus gB glyco
	7	38	6.4	1037	21	AAA59242	Exons E, C and A o
	8	38	6.4	1159	21	AAA59240	An EcoRI fragment
	9	38	6.4	1472	21	AAA59241	Exons D, C, B and
	10	37.6	6.3	2115	20	AAZ32263	Human dunce-like p
	11	37.6	6.3	2115	21	AAA88198	pPDE39 human dunce
	12	37.6	6.3	72750	21	AAA81468	N. meningitidis pa
c	13	37.6	6.3	349980	21	AAF21544	Neisseria meningit
	14	37.2	6.3	921	20	AAX33558	Rice beta-glucanas
	15	37.2	6.3	996	20	AAX33568	Rice Gns6 CDS. Or
	16	37.2	6.3	2169	20	AAX33534	Rice beta-glucanas
	17	36.8	6.2	3737	21	AAA64656	Partial sequence c
	18	36.8	6.2	3776	17	AAT42903	TRP-1 protein codi
c	19	36.6	6.2	203	19	AAV30271	Glutamine rich reg
c	20	36.6	6.2	203	19	AAV17226	SCA2 gene CAG repe
	21	36.6	6.2	913	17	AAT30253	Cotton fibre cell-
	22	36.6	6.2	913	17	AAT13034	Cotton fibre-speci
	23	36.6	6.2	913	17	AAT07199	Cotton fiber-speci
	24	36.6	6.2	913	18	AAT70041	Cotton fibre speci
	25	36.6	6.2	913	18	AAT62610	Cotton fibre speci
	26	36.6	6.2	913	21	AAZ35545	cDNA sequence a co
	27	36.6	6.2	1966	20	AAX61220	Mouse DNA demethyl
	28	36.6	6.2	1984	17	AAT13030	Cotton fibre-speci
	29	36.6	6.2	1985	17	AAT30250	Cotton fibre clone
	30	36.6	6.2	1985	18	AAT70036	Cotton H6 gene and
c	31	36.6	6.2	3324	20	AAX60262	Nucleic acid seque
c	32	36.4	6.1	2415	18	AAT86757	cDNA of the M3/6 g
c	33	36.4	6.1	2453	18	AAT86758	cDNA of the M3/6 g
c	34	36.2	6.1	633	11	AAQ03322	Genomic Eimeria te
	35	36.2	6.1	862	21	AAF09682	Fusarium venenatum
	36	36.2	6.1	3201	21	AAA14955	DNA encoding a PDE
c	37	36	6.1	795	19	AAV55830	FLGA insert stabil
c	38	36	6.1	799	19	AAV55831	Nucleotide sequenc
	39	36	6.1	1448	21	AAF21931	Human breast and o
c	40	36	6.1	1925	20	AAX90924	Epstein Barr Virus
c	41	36	6.1	1926	21	AAA50254	Epstein Barr virus
c	42	36	6.1	2580	21	AAA75454	Nucleotide sequenc
	43	36	6.1	5452	20	AAX90923	Anti-sense strand
	44	36	6.1	8705	20	AAZ23778	Vector pShuttle DN
c	45	36	6.1	9600	19	AAV21683	Vector plasmid pCM

ALIGNMENTS

RESULT 1
AAX84092
ID AAX84092 standard; cDNA; 3402 BP.
XX
AC AAX84092;
XX
DT 27-AUG-1999 (first entry)
XX
DE T. cruzi PTc40 coding sequence.
XX
KW PTc40; Tc40; infection; diagnosis; immune complex; antigenic determinant;
KW therapy; antibody; ds.
XX
OS Trypanosoma cruzi.
XX
PN WO9929867-A1.
XX
PD 17-JUN-1999.
XX
PF 10-DEC-1998; 98WO-IB01987.
XX
PR 10-DEC-1997; 97US-0988242.
XX
PA (INMR) BIO MERIEUX.
XX
PI Jolivet M, Lesenechal M, Mandrand B, Paranhos-Baccala G;
XX
DR WPI; 1999-394978/33.
DR P-PSDB; AAY22124.
XX
PT New Trypanosoma cruzi antigen
XX
PS Claim 1; Page 52-56; 65pp; English.
XX

```
CC   This sequence encodes the Trypanosoma cruzi PTc40 protein of the
CC   invention, and is designated Tc40. The PTc40 antigenic determinant is
CC   useful as a reagent for detection and/or monitoring of Trypanosoma cruzi
CC   infection from samples including blood serum or plasma, urine, saliva, or
CC   tears, by contacting with the sample and detecting an immune complex. The
CC   PTc40 antigenic determinant, the vector, expression cassette, cell or
CC   antibody are useful for treatment or prevention (vaccine) of a
CC   Trypanosoma cruzi infection in a man or animal. Current Trypanosoma cruzi
CC   antigens are obtained from protein fractions of the noninfectious stage
CC   of the parasite, and these do not allow sufficient production of antigens
CC   for use in reliable serological diagnostic tests. The strain to strain
CC   polymorphism reduces reliability of the tests.
XX
SQ   Sequence 3402 BP; 888 A; 821 C; 956 G; 737 T; 0 other;


  Query Match             100.0%;  Score 594;  DB 20;  Length 3402;
  Best Local Similarity   100.0%;  Pred. No. 8.3e-146;
  Matches  594;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 caggtacagcgtaacggcttttgcttcaatcgtacagccgacggtagctgcgtcctggct 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1232 caggtacagcgtaacggcttttgcttcaatcgtacagccgacggtagctgcgtcctggct 1291

Qy      61 gacatgtcgattcgattgacgatcttccatctccggtcctcccgcagggaagaacagcag 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1292 gacatgtcgattcgattgacgatcttccatctccggtcctcccgcagggaagaacagcag 1351

Qy     121 ccaggccaaaaaacatcggtagtggcgacggcgaaaccggggtgtgtgtcctcgggcact 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1352 ccaggccaaaaaacatcggtagtggcgacggcgaaaccggggtgtgtgtcctcgggcact 1411

Qy     181 gacgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatcccctgcatca 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1412 gacgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatcccctgcatca 1471

Qy     241 ccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggctgag 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1472 ccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggctgag 1531

Qy     301 ccgcacgtggggagcaagatcattgctaatctagtgaatcagctggggattaatgtcacc 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1532 ccgcacgtggggagcaagatcattgctaatctagtgaatcagctggggattaatgtcacc 1591

Qy     361 caaaggagcgtcgtcagcactggagcgccggccacgacgaggtctacggcggtgacgtcc 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1592 caaaggagcgtcgtcagcactggagcgccggccacgacgaggtctacggcggtgacgtcc 1651

Qy     421 acgactaccgccccgcagcgaacaagtccatacgggcacaatggccgacctgtgacggct 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1652 acgactaccgccccgcagcgaacaagtccatacgggcacaatggccgacctgtgacggct 1711

Qy     481 ggattggtggcagctaatagtggtgccagcgcggcctcgtctcccacagccgcggcgaaa 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1712 ggattggtggcagctaatagtggtgccagcgcggcctcgtctcccacagccgcggcgaaa 1771

Qy     541 ccaacaggagaagaaaaggcctccgcggcatgtgaaacgagctccgtggcgata 594
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1772 ccaacaggagaagaaaaggcctccgcggcatgtgaaacgagctccgtggcgata 1825


RESULT   2
AAT27310
ID   AAT27310 standard; cDNA; 3402 BP.
XX
AC   AAT27310;
XX
DT   26-NOV-1996  (first entry)
XX
DE   Trypanosoma cruzi epimastigotic PTc100t antigen gene.
XX
KW   Antigen; Trypanosoma cruzi; epimastigote; serum; Chagas disease; probe;
KW   Primer; PCR; polymerase chain reaction; amplification; antibody; ds.
XX
OS   Trypanosoma cruzi.
XX
FH   Key             Location/Qualifiers
FT   CDS             266..3013
FT                   /*tag= a
FT                   /product= PTc100t epimastogotic antigen
XX
PN   FR2723589-A1.
XX
PD   16-FEB-1996.
XX
PF   12-AUG-1994;   94FR-0010132.
XX
PR   12-AUG-1994;   94FR-0010132.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Jolivet M,  Lesenechal M,  Paranhos-Baccala G;
XX
DR   WPI; 1996-190287/20.
DR   P-PSDB; AAR91615.
XX
PT   New nucleic acid encoding Trypanosoma cruzi epimastigotic antigen -
PT   useful for diagnosis, monitoring and therapy of Chagas disease
XX
PS   Claim 1; Page 24-26; 55pp; French.
XX
CC   This is the nucleotide sequence encoding a novel isolated antigenic
CC   protein from Trypanosoma cruzi epimastigotes, designated PTc100t.
CC   The clone Tc50 was isolated from a T.cruzi genomic expression library in
CC   lambda gt11, using a mixture of sera from patients with Chagas disease.
CC   Clone Tc50 contained an 594 bp insert corresp. to nucleotides 1232-1825
CC   of this sequence.  The Tc50 sequence was subsequently used to probe a
CC   Southern blot of restriction enzyme digested T.cruzi DNA and also screen
CC   a lambda gt10 library to isolate a 1041 bp EcoRI fragment corresp. to
CC   nucleotides 1403-2443 of PTc100t.  Primers (AAT27311-5) were synthesised
CC   based on the sequences of the 594 and 1041 bp fragments and used to
CC   amplify the PTc100t clone as 3 fragments from cDNA derived from mRNA
CC   purified from T.cruzi epimastigotes.  The protein or antibodies raised
CC   against it can be used in the detection and monitoring of T.cruzi
CC   infection i.e. Chagas disease.
XX
SQ   Sequence 3402 BP; 889 A; 818 C; 958 G; 737 T; 0 other;


  Query Match             99.5%;  Score 590.8;  DB 17;  Length 3402;
  Best Local Similarity   99.7%;  Pred. No. 5.7e-145;
  Matches  592;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 caggtacagcgtaacggcttttgcttcaatcgtacagccgacggtagctgcgtcctggct 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1232 caggtacagcgtaacggcttttgcttcaatcgtacagccgacggtagctgcgtcctggct 1291

Qy      61 gacatgtcgattcgattgacgatcttccatctccggtcctcccgcagggaagaacagcag 120
          |||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1292 gacatgtcgaatcgattgacgatcttccatctccggtcctcccgcagggaagaacagcag 1351

Qy     121 ccaggccaaaaaacatcggtagtggcgacggcgaaaccggggtgtgtgtcctcgggcact 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1352 ccaggccaaaaaacatcggtagtggcgacggcgaaaccggggtgtgtgtcctcgggcact 1411

Qy     181 gacgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatcccctgcatca 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1412 gacgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatcccctgcatca 1471

Qy     241 ccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggctgag 300
          ||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db    1472 ccccctgtttcagcgccagcgaaggcagccgcgcctcctgccgcggcgcgatcggctgag 1531

Qy     301 ccgcacgtggggagcaagatcattgctaatctagtgaatcagctggggattaatgtcacc 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1532 ccgcacgtggggagcaagatcattgctaatctagtgaatcagctggggattaatgtcacc 1591
```

```
Qy     361 caaaggagcgtcgtcagcactggagcgccggccacgacgaggtctacggcggtgacgtcc 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1592 caaaggagcgtcgtcagcactggagcgccggccacgacgaggtctacggcggtgacgtcc 1651

Qy     421 acgactaccgccccgcagcgaacaagtccatacgggcacaatggccgacctgtgacggct 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1652 acgactaccgccccgcagcgaacaagtccatacgggcacaatggccgacctgtgacggct 1711

Qy     481 ggattggtggcagctaatagtggtgccagcgcggcctcgtctcccacagccgcggcgaaa 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1712 ggattggtggcagctaatagtggtgccagcgcggcctcgtctcccacagccgcggcgaaa 1771

Qy     541 ccaacaggagaagaaaaggcctccgcggcatgtgaaacgagctccgtggcgata 594
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1772 ccaacaggagaagaaaaggcctccgcggcatgtgaaacgagctccgtggcgata 1825
```

RESULT 3
AAZ44307/c
ID AAZ44307 standard; DNA; 477 BP.
XX
AC AAZ44307;
XX
DT 04-APR-2000 (first entry)
XX
DE Human SCA7 genomic DNA fragment 1.
XX
KW SCA7; human; spinocerebellar ataxia type 7; SCA1; SCA2; SCA3; SCA6;
KW repeat expansion detection; RED analysis; detection; ds.
XX
OS Homo sapiens.
XX
PN CA2245310-A.
XX
PD 19-FEB-1999.
XX
PF 19-AUG-1998; 98CA-2245310.
XX
PR 19-AUG-1997; 97US-0056170.
XX
PA (MINU) UNIV MINNESOTA.
XX
PI Koob MD, Ranum LP;
XX
DR WPI; 2000-098181/09.
XX
PT Identifying individuals at risk of developing spinocerebellar ataxia
PT type 7 by analyzing trinucleotide repeat regions of spinocerebellar
PT ataxia type 7 gene -
XX
PS Claim 36; Page 41; 66pp; English.
XX
CC This invention describes a novel method for identifying individuals at
CC risk for developing spinocerebellar ataxia type 7 (SCA7). The method
CC comprises analyzing the CAG repeat region of a SCA7 gene to detect CAG
CC repeats, where individuals at risk have at least 30 CAG repeats and
CC those not at risk have less than 19 CAG repeats. The method is useful
CC for identifying individuals at risk of developing SCA7 and also those
CC at risk of developing SCA1, 2, 3 or 6. The use of genomic DNA in the
CC repeat expansion detection (RED) analysis allows isolation of any
CC potential trinucleotide repeat expansion regardless of the expression
CC pattern. Utilization of different oligonucleotides in the RED assay
CC allows any of the possible trinucleotide repeats to be detected, and
CC the cycled nature of the reaction makes it extremely sensitive. This
CC sequence represents a genomic DNA fragment of the human SCA7 gene
CC which is described in the method of the invention.
XX
SQ Sequence 477 BP; 99 A; 164 C; 172 G; 42 T; 0 other;

Query Match 6.8%; Score 40.4; DB 21; Length 477;
Best Local Similarity 53.9%; Pred. No. 0.15;

Matches 83; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

```
Qy     145 gcgacggcgaaaccggggtgtgtgtcctcgggcactgacgcggcgagtagcagtcatacc 204
           | | |||||    ||||   ||   || | ||  |||  | ||||        |  | |
Db     334 GTGGCGGCGGTGGCGGGTGCTGCTGCCGCTGGGGCTGCGGAGGCGGCGGCTGCTGCTGCT 275

Qy     205 aatacgacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaag 264
             | |  || |||| ||||||||  |  ||||  |  |  |||| | | ||  | ||
Db     274 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 215

Qy     265 gcagccgcgcctcctgccgcggcgcgatcggctg 298
           || || ||  || |||| || ||     | ||||
Db     214 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 181
```

RESULT 4
AAA29550/c
ID AAA29550 standard; DNA; 567 BP.
XX
AC AAA29550;
XX
DT 09-AUG-2000 (first entry)
XX
DE HIV codon altered env nucleotide sequence #2.
XX
KW Erythropoietin; EPO; G-CSF; granulocyte colony stimulating factor;
KW wobble; codon altered gene; shuffling; modification; vaccine;
KW insulin; peptide hormone; growth factor; cytokine; interferon;
KW interleukin; leukaemia inhibitory factor; oncostatin M;
KW transcription activator; expression activator; infectious organism; ds.
XX
OS Human immunodeficiency virus type 1.
OS Synthetic.
XX
PN WO200018906-A2.
XX
PD 06-APR-2000.
XX
PF 28-SEP-1999; 99WO-US22588.
XX
PR 29-SEP-1998; 98US-0102362.
PR 29-JAN-1999; 99US-0117729.
PR 05-FEB-1999; 99US-0118813.
PR 24-JUN-1999; 99US-0141049.
XX
PA (MAXY-) MAXYGEN INC.
XX
PI Patten PA, Liu L, Stemmer WPC;
XX
DR WPI; 2000-303449/26.
XX
PT Novel methods for recombining codon-altered libraries of nucleic acids
PT used to produce new proteins and new vectors with reduced rates of
PT reversion to wild type -
XX
PS Example; Fig 18A; 92pp; English.
XX
CC A method has been developed of making codon altered nucleic acids (NAs),
CC comprising providing a NA sequence (NA1) which encodes a polypeptide
CC (P1), providing codon altered NA sequences, each encoding P1 or a
CC modified form of it, and recombining the codon altered NA sequences to
CC produce a target codon altered NA which encodes a second protein.
CC The method of the invention can be used for recombining codon-altered
CC libraries of nucleic acids to produce new proteins, which have
CC improvements in a desirable characteristic. Target nucleic acids
CC include those coding for therapeutic proteins such as erythropoietin
CC (EPO), insulin, peptide hormones, growth factors, cytokines, interferons,
CC interleukins, leukaemia inhibitory factor, and oncostatin M, as well as
CC transcription and expression activators and proteins from infectious
CC organisms for use as vaccines. The method can also be used to produce
CC attenuated viruses which have reduced rates of reversion to wild type.
CC The present sequence represents an HIV codon altered env nucleotide
CC sequence, which is used in an example from the present invention.

```
XX
SQ   Sequence 567 BP; 32 A; 129 C; 222 G; 184 T; 0 other;


  Query Match             6.7%;  Score 39.8;  DB 21;  Length 567;
  Best Local Similarity   50.3%;  Pred. No. 0.23;
  Matches   98;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy     386 cgccggccacgacgaggtctacggcggtgacgtccacgactaccgccccgcagcgaacaa 445
           || || | ||||||| | | ||| ||  |||| | ||||| ||  |  ||  |   || |
Db     477 CGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGA 418

Qy     446 gtccatacgggcacaatggccgacctgtgacggctggattggtggcagctaatagtggtg 505
           |      |   | |    | | |   | || | |   |  |||||       |  |
Db     417 CGACGACGACGACGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 358

Qy     506 ccagcgcggcctcgtctcccacagccgcggcgaaaccaacaggagaagaaaaggcctccg 565
           |    || ||  |  |  |  |||| || ||   | || ||| || || |   ||  | |
Db     357 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 298

Qy     566 cggcatgtgaaacga 580
           | |||   | ||| |
Db     297 CAGCAGCAGCAACAA 283


RESULT   5
AAV18480
ID   AAV18480 standard; cDNA to mRNA; 2790 BP.
XX
AC   AAV18480;
XX
DT   18-AUG-1998  (first entry)
XX
DE   BOP1 cDNA.
XX
KW   Corticotroph; pituitary; BOP1; tumour suppressor; constitutive;
KW   inducible; alzheimer's disease; nuclear transcription factor; apoptosis;
KW   cell cycle; neuronal disorder; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             542..2545
FT                   /*tag=  a
FT                   /product= "BOP1 protein"
XX
PN   WO9813489-A1.
XX
PD   02-APR-1998.
XX
PF   22-SEP-1997;   97WO-EP05198.
XX
PR   23-SEP-1996;   96US-0718661.
XX
PA   (CNRS ) CENT NAT RECH SCI.
PA   (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI   Journot L,  Spengler D;
XX
DR   WPI; 1998-230701/20.
DR   P-PSDB; AAW48760.
XX
PT   New isolated tumour suppressor gene - useful for developing products
PT   for use in diagnosis and treatment of tumour(s) or neuronal
PT   disorder(s)
XX
PS   Claim 1; Pages 72-76; 118pp; English.
XX
CC   The present sequence represents the BOP1 cDNA isolated from the
CC   mouse corticotroph pituitary tumour cell line AtT-20 cDNA library.
CC   The protein encoded by the BOP1 cDNA displays a tumour suppressing
CC   activity when it was constitutively and inducibly expressed in
CC   tumour cells.  The BOP1 cDNA and the protein it encodes are claimed
CC   to be useful in the preparation of therapeutic compositions, useful
CC   for treating, preventing or delaying the recurrence of a tumour or
CC   neuronal disorders, e.g. genetic diseases or acquired degenerative
CC   encephalopathies such as Alzheimer's disease.  The BOP1 protein is
CC   also claimed to be able to induce apoptosis resulting in inhibition
CC   of tumour cell growth, to suppress tumour formation, to induce G1
CC   arrest of the cell cycle and to act as nuclear transcription factor.
XX
SQ   Sequence 2790 BP; 667 A; 783 C; 714 G; 626 T; 0 other;


  Query Match             6.6%;  Score 39.2;  DB 19;  Length 2790;
  Best Local Similarity   54.1%;  Pred. No. 0.51;
  Matches   80;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy     137 cggtagtggcgacggcgaaaccggggtgtgtgtcctcgggcactgacgcggcgagtagca 196
           | | ||  || || || |   | |         || | | ||||| | | || | |  ||
Db    2114 cagcagcagccaccgccactgccgccgccaccaccgctgccactgccacagccactgcca 2173

Qy     197 gtcataccaatacgacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgc 256
           ||   ||    || ||  | ||| | | |||| || | |  |||    | | | |||| |
Db    2174 gtgccacagccactaccacagccacagatgcagccacagtttcagttgcagatccagccc 2233

Qy     257 cagccaaggcagccgcgcctcctgccgc 284
           |||     ||  || |  || |||||||
Db    2234 cagatgcagctaccacagctgctgccgc 2261


RESULT   6
AAT16480/c
ID   AAT16480 standard; DNA; 2943 BP.
XX
AC   AAT16480;
XX
DT   11-MAY-1996  (first entry)
XX
DE   SA8 virus gB glycoprotein coding sequence.
XX
KW   Herpes simian monkey SA8 virus gB glycoprotein; immunoassay;
KW   diagnosis; herpes B virus; ss.
XX
OS   Herpes simian monkey SA8 virus.
XX
FH   Key             Location/Qualifiers
FT   CDS             87..2744
FT                   /*tag=  a
FT                   /product= gB_glycoprotein
XX
PN   US5487969-A.
XX
PD   30-JAN-1996.
XX
PF   01-APR-1993;   93US-0042747.
XX
PR   01-APR-1993;   93US-0042747.
XX
PA   (SWBI-) SOUTHWEST FOUND BIOMEDICAL RES.
XX
PI   Black D,  Eberle R,  Hilliard J,  Scinicariello F;
XX
DR   WPI; 1996-105220/11.
DR   P-PSDB; AAR92747.
XX
PT   Detection of herpes B virus by PCR amplification of sample DNA - to
PT   detect a specific herpes simian monkey B virus DNA segment.
XX
PS   Disclosure; Column 23-30; 22pp; English.
XX
CC   This sequence encodes the herpes simian monkey SA8 virus gB
CC   glycoprotein.  Such sequences have been used to develop synthetic
CC   DNA primers which have homologous sequences of conserved regions
CC   which flank a divergent region of the gB glycoprotein gene.  In
CC   addition, proteins such as the gB glycoprotein have potential use in
```

```
CC   the development of serological immunoassays.  One approach is to
CC   synthesize peptides which, based on the properties of the predicted
CC   protein sequence, are likely to be immunologically active.  Such
CC   peptides can be used as substrate antigens in immunoassays to detect
CC   serum antibodies which recognize this specific peptide sequence.
CC   Synthetic peptides may also be used to produce antibodies against
CC   specific regions of the gB glycoprotein which are unique to one
CC   virus.  These can then be used to develop virus-specific
CC   immunoassays for differentiation of SA8 virus from other primate
CC   alpha-herpes viruses and for identification of antibodies directed
CC   against SA8 virus in primate serum samples.
XX
SQ   Sequence 2943 BP; 462 A; 1103 C; 983 G; 395 T; 0 other;


  Query Match             6.4%;  Score 38.2;  DB 17;  Length 2943;
  Best Local Similarity   47.7%;  Pred. No. 0.94;
  Matches  112;  Conservative    0;  Mismatches  123;  Indels    0;  Gaps    0;

Qy    209 cgacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaaggcag 268
          || | ||  |||   |  | ||  |  |   |  | | |     ||| | |    ||| |
Db    255 CGTCCTCGTCCGGGACCTCCTCGTCCTCGGGAAGCGCGGGCGTGGCGGCGGGGTCGGCGG 196

Qy    269 ccgcgcctcctgccgcggcgcgatcggctgagccgcacgtggggagcaagatcattgcta 328
           || | |  | ||||| ||| |  |||||   |||  || || || || |   |  || |
Db    195 TCGGGGCCGCCGCCGCCGCGGGGGCGGCTCGTCCGGCCGCGGCGATCACGGCGAGGGCGA 136

Qy    329 atctagtgaatcagctggggattaatgtcacccaaaggagcgtcgtcagcactggagcgc 388
              |     |  | |   | |          | | | |  || | | ||
Db    135 GGAGGACGGGGAGGGGAAGAAAGGAGGGGGGGGTGCCGCGAGGCCGCATCCCGGGGCGCG 76

Qy    389 cggccacgacgaggtctacggcggtgacgtccacgactaccgccccgcagcgaac 443
          ||| | | ||| || | ||  ||  | |||| ||  |  ||  || || |||| |
Db     75 CGGGCGCTACGCGGGCGACTACGCGGGCGTCGACTCCGGCCCGCCGGCCGCGAGC 21


RESULT   7
AAA59242
ID   AAA59242 standard; cDNA; 1037 BP.
XX
AC   AAA59242;
XX
DT   07-NOV-2000  (first entry)
XX
DE   Exons E, C and A of the SCA8 repeat region.
XX
KW   Repeat region; spinocerebellar ataxia type 8; SCA8; long arm;
KW   chromosome 13; polymorphic CTA repeat; CTG repeat; ss.
XX
OS   Homo sapiens.
XX
PN   CA2283758-A1.
XX
PD   28-APR-2000.
XX
PF   27-SEP-1999;   99CA-2283758.
XX
PR   28-OCT-1998;   98US-0181585.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Benzow KA,  Moseley-alldredge ML,  Ranum LPW,  Koob MD;
XX
DR   WPI; 2000-491456/44.
XX
PT   Novel nucleic acid molecule containing repeat region of spinocerebellar
PT   ataxia type 8 coding sequence, useful as primer or probe for diagnosing
PT   or determining risk for developing spinocerebellar ataxia -
XX
PS   Claim 25; Fig 7C; 68pp; English.
XX
CC   The present sequence represents a nucleic acid molecule containing a
CC   repeat region of a spinocerebellar ataxia type 8 (SCA8) coding
CC   sequence which is located within the long arm of chromosome 13. The
CC   SCA8 coding sequence contains polymorphic CTA repeats and CTG repeats.
CC   Generally, the CTA repeat is unstable, and can vary between individuals
CC   in different families (the number of CTA repeats in the region does not
CC   vary between individuals of a family). Fragments of the nucleic acid
CC   molecule are useful as primers and probes for determining whether an
CC   individual has or is at risk of developing spinocerebellar ataxia 8.
XX
SQ   Sequence 1037 BP; 229 A; 261 C; 242 G; 305 T; 0 other;


  Query Match             6.4%;  Score 38;  DB 21;  Length 1037;
  Best Local Similarity   56.3%;  Pred. No. 0.8;
  Matches   71;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy    173 cgggcactgacgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatccc 232
          | |||  |    |  | | ||  | |  ||| | ||| ||| |||| |||||||  |
Db    629 ctggctttactactactactactactactactactactactgctactgctgctgctg 688

Qy    233 ctgcatcacccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgat 292
          ||||  |  |  ||| | | ||  | ||    || || ||  || |||| || ||
Db    689 ctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 748

Qy    293 cggctg 298
          | ||||
Db    749 ctgctg 754


RESULT   8
AAA59240
ID   AAA59240 standard; DNA; 1159 BP.
XX
AC   AAA59240;
XX
DT   07-NOV-2000  (first entry)
XX
DE   An EcoRI fragment that includes SCA8 repeat region.
XX
KW   Repeat region; spinocerebellar ataxia type 8; SCA8; long arm;
KW   chromosome 13; polymorphic CTA repeat; CTG repeat; ss.
XX
OS   Homo sapiens.
XX
PN   CA2283758-A1.
XX
PD   28-APR-2000.
XX
PF   27-SEP-1999;   99CA-2283758.
XX
PR   28-OCT-1998;   98US-0181585.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Benzow KA,  Moseley-alldredge ML,  Ranum LPW,  Koob MD;
XX
DR   WPI; 2000-491456/44.
XX
PT   Novel nucleic acid molecule containing repeat region of spinocerebellar
PT   ataxia type 8 coding sequence, useful as primer or probe for diagnosing
PT   or determining risk for developing spinocerebellar ataxia -
XX
PS   Claim 26; Fig 7A; 68pp; English.
XX
CC   The present sequence represents a nucleic acid molecule containing a
CC   repeat region of a spinocerebellar ataxia type 8 (SCA8) coding
CC   sequence which is located within the long arm of chromosome 13. The
CC   SCA8 coding sequence contains polymorphic CTA repeats and CTG repeats.
CC   Generally, the CTA repeat is unstable, and can vary between individuals
CC   in different families (the number of CTA repeats in the region does not
CC   vary between individuals of a family). Fragments of the nucleic acid
CC   molecule are useful as primers and probes for determining whether an
CC   individual has or is at risk of developing spinocerebellar ataxia 8.
XX
```

```
SQ   Sequence 1159 BP; 263 A; 263 C; 196 G; 437 T; 0 other;

  Query Match             6.4%;  Score 38;  DB 21;  Length 1159;
  Best Local Similarity   56.3%;  Pred. No. 0.83;
  Matches   71;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy    173 cgggcactgacgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatccc 232
          | |||  |    |  | | ||  | |  ||| | ||| ||| |||| |||||||  |
Db    441 ctggctttactactactactactactactactactactactactgctactgctgctgctg 500

Qy    233 ctgcatcacccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgat 292
          ||||  |  |  ||| | | ||  | ||    || || ||  || |||| || ||
Db    501 ctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 560

Qy    293 cggctg 298
          | ||||
Db    561 ctgctg 566


RESULT   9
AAA59241
ID   AAA59241 standard; cDNA; 1472 BP.
XX
AC   AAA59241;
XX
DT   07-NOV-2000  (first entry)
XX
DE   Exons D, C, B and A of the SCA8 repeat region.
XX
KW   Repeat region; spinocerebellar ataxia type 8; SCA8; long arm;
KW   chromosome 13; polymorphic CTA repeat; CTG repeat; ss.
XX
OS   Homo sapiens.
XX
PN   CA2283758-A1.
XX
PD   28-APR-2000.
XX
PF   27-SEP-1999;   99CA-2283758.
XX
PR   28-OCT-1998;   98US-0181585.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Benzow KA,  Moseley-alldredge ML,  Ranum LPW,  Koob MD;
XX
DR   WPI; 2000-491456/44.
XX
PT   Novel nucleic acid molecule containing repeat region of spinocerebellar
PT   ataxia type 8 coding sequence, useful as primer or probe for diagnosing
PT   or determining risk for developing spinocerebellar ataxia -
XX
PS   Claim 24; Fig 7B; 68pp; English.
XX
CC   The present sequence represents a nucleic acid molecule containing a
CC   repeat region of a spinocerebellar ataxia type 8 (SCA8) coding
CC   sequence which is located within the long arm of chromosome 13. The
CC   SCA8 coding sequence contains polymorphic CTA repeats and CTG repeats.
CC   Generally, the CTA repeat is unstable, and can vary between individuals
CC   in different families (the number of CTA repeats in the region does not
CC   vary between individuals of a family). Fragments of the nucleic acid
CC   molecule are useful as primers and probes for determining whether an
CC   individual has or is at risk of developing spinocerebellar ataxia 8.
XX
SQ   Sequence 1472 BP; 337 A; 368 C; 378 G; 388 T; 1 other;

  Query Match             6.4%;  Score 38;  DB 21;  Length 1472;
  Best Local Similarity   56.3%;  Pred. No. 0.88;
  Matches   71;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy    173 cgggcactgacgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatccc 232
          | |||  |    |  | | ||  | |  ||| | ||| ||| |||| |||||||  |
Db   1064 ctggctttactactactactactactactactactactactactgctactgctgctgctg 1123

Qy    233 ctgcatcacccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgat 292
          ||||  |  |  ||| | | ||  | ||    || || ||  || |||| || ||
Db   1124 ctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1183

Qy    293 cggctg 298
          | ||||
Db   1184 ctgctg 1189


RESULT  10
AAZ32263
ID   AAZ32263 standard; cDNA; 2115 BP.
XX
AC   AAZ32263;
XX
DT   19-JAN-2000  (first entry)
XX
DE   Human dunce-like phosphodiesterase pPDE39 encoding cDNA.
XX
KW   Phosphodiesterase; dunce-like phosphodiesterase; PDE; DPD; cAMP;
KW   RAS-related protein; immunoreactive; detection; genetic defect;
KW   bronchodilation; increased myocardial contractility;
KW   anti-inflammation; ss.
XX
OS   Homo sapiens.
XX
PN   US5977305-A.
XX
PD   02-NOV-1999.
XX
PF   07-JUN-1995;   95US-0474379.
XX
PR   01-MAR-1994;   94US-0206188.
PR   20-APR-1990;   90US-0511715.
PR   19-APR-1991;   91US-0688352.
XX
PA   (COLD-) COLD SPRING HARBOR LAB.
XX
PI   Colicelli JJ,  Wigler MH;
XX
DR   WPI; 1999-619709/53.
DR   P-PSDB; AAY49826.
XX
PT   New isolated RAS-related polypeptides and mammalian cyclic nucleotide
PT   phosphodiesterases, used for screening for agents which can modify
PT   complement or suppress genetic defects  -
XX
PS   Example 6; Column 165-170; 145pp; English.
XX
CC   The present invention describes new isolated RAS-related polypeptides
CC   and mammalian cyclic nucleotide phosphodiesterases (PDEs). RAS-related
CC   polypeptides are capable of complementing a defective RAS function in
CC   yeast. The products can be used for screening for agents which can
CC   modify, complement or suppress a genetic defect in a biochemical
CC   pathway in which cAMP participates, or in a biochemical pathway which
CC   is controlled, directly or indirectly, by a RAS protein and other
CC   proteins affecting cell growth and maintenance. Developing agents that
CC   will selectively act upon PDEs is directed toward reproducing the
CC   desirable effects of cyclic nucleotides, e.g. bronchodilation,
CC   increased myocardial contractility, anti-inflammation, yet without
CC   causing the undesirable effects, e.g. increased heart rate or enhanced
CC   lipolysis. The products can also be used for therapeutic, diagnostic
CC   and prognostic uses. AAZ32229 to AAZ32285, and AAY49803 to AAY49830,
CC   represent sequences used in the exemplification of the present
CC   invention.
XX
SQ   Sequence 2115 BP; 536 A; 605 C; 501 G; 473 T; 0 other;

  Query Match             6.3%;  Score 37.6;  DB 20;  Length 2115;
```

```
  Best Local Similarity   63.0%;  Pred. No. 1.2;
  Matches   58;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy     212 cttctgccgctgctgcatcccctgcatcacccctgtttcagcgccagccaaggcagccg 271
           || |  |||| || ||  |||  ||  || |||| |   ||| | |||||   || ||||
Db     577 ctgcccccgccgccgccaccctcgccccagccccagccccagtgtcagccgccgccgccg 636

Qy     272 cgcctcctgccgcggcgcgatcggctgagccg 303
           | || ||||||||| || |   | || |  | |
Db     637 ccccccctgccgccgccccgccgccgcccgg 668


RESULT  11
AAA88198
ID   AAA88198 standard; cDNA; 2115 BP.
XX
AC   AAA88198;
XX
DT   14-DEC-2000  (first entry)
XX
DE   pPDE39 human dunce-like PDE nucleotide sequence SEQ ID NO:60.
XX
KW   Detection; mammalian gene; yeast; microorganism; identification;
KW   phenotype; characteristic; dunce-like phosphodiesterase; PDE; RAS;
KW   RAS-related protein; genetic defect; hybridisation; probe; ss.
XX
OS   Homo sapiens.
OS   Saccharomyces cerevisiae.
XX
PN   US6100025-A.
XX
PD   08-AUG-2000.
XX
PF   01-MAR-1994;   94US-0206188.
XX
PR   20-APR-1990;   90US-0511715.
PR   19-APR-1991;   91US-0688352.
XX
PA   (COLD-) COLD SPRING HARBOR LAB.
XX
PI   Colicelli JJ,  Wigler MH;
XX
DR   WPI; 2000-531664/48.
DR   P-PSDB; AAB20637.
XX
PT   Novel isolated DNA encoding a mammalian cyclic nucleotide
PT   phosphodiesterase is present in plasmids pPDE46, pPDE43 or pPDE339 and
PT   is used to modify a genetic defect in a biochemical pathway in which
PT   cAMP participates  -
XX
PS   Claim 2; Column 171-176; 145pp; English.
XX
CC   The present invention describes a purified and isolated DNA (I) which
CC   encodes a mammalian cyclic nucleotide phosphodiesterase and is an insert
CC   present in the plasmids pPDE46 (ATCC 69552), pPDE43 (ATCC 69551) or
CC   pPDE339 (ATCC 69550). The DNA molecules are used to modify, complement
CC   or suppress a genetic defect in a biochemical pathway in which cAMP
CC   participates and are also used as hybridisation probes. The present
CC   invention also describes methods for detecting mammalian genes encoding
CC   proteins which can function in microorganisms, particularly yeast, to
CC   modify, complement, or suppress a genetic defect associated with an
CC   identifiable phenotypic alteration or characteristic in the
CC   microorganism. AAA88162 to AAA88218 and AAB29614 to AAB20640 represent
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence 2115 BP; 536 A; 605 C; 501 G; 473 T; 0 other;


  Query Match             6.3%;  Score 37.6;  DB 21;  Length 2115;
  Best Local Similarity   63.0%;  Pred. No. 1.2;
  Matches   58;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy     212 cttctgccgctgctgcatcccctgcatcacccctgtttcagcgccagccaaggcagccg 271
           || |  |||| || ||  |||  ||  || |||| |   ||| | |||||   || ||||
Db     577 ctgcccccgccgccgccaccctcgccccagccccagccccagtgtcagccgccgccgccg 636

Qy     272 cgcctcctgccgcggcgcgatcggctgagccg 303
           | || ||||||||| || |   | || |  | |
Db     637 ccccccctgccgccgccccgccgccgcccgg 668


RESULT  12
AAA81468
ID   AAA81468 standard; DNA; 72750 BP.
XX
AC   AAA81468;
XX
DT   04-DEC-2000  (first entry)
XX
DE   N. meningitidis partial DNA sequence gnm_16 SEQ ID NO:16.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; genome; immunogenic;
KW   antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW   Meningococcus B; MenB; ds.
XX
OS   Neisseria meningitidis.
XX
PN   WO200022430-A2.
XX
PD   20-APR-2000.
XX
PF   08-OCT-1999;   99WO-US23573.
XX
PR   09-OCT-1998;   98US-0103794.
PR   30-APR-1999;   99US-0132068.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Frazer CM,  Hickey E,  Peterson J,  Tettelin H,  Venter JC;
PI   Masignani V,  Galeotti C,  Mora M,  Ratti G,  Scarselli M,  Scarlato V;
PI   Rappuoli R,  Pizza M;
XX
DR   WPI; 2000-318079/27.
XX
PT   Isolated nucleotide sequences of Neisseria meningitidis which can be
PT   used in the diagnosis and treatment of N. meningitidis infection and
PT   other Neisserial infections, for example, N.gonorrhoea -
XX
PS   Claim 7; Page 406-428; 1760pp; English.
XX
CC   The present invention describes methods of obtaining immunogenic
CC   proteins from Neisseria genomic sequences. AAA81453 to AAA82414
CC   represent specifically claimed Neisseria meningitidis genomic DNA
CC   sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
CC   Neisseria DNA sequences and their corresponding proteins; AAA81254 to
CC   AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
CC   isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
CC   AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
CC   sequences, which are all used in the exemplification of the present
CC   invention. The nucleic acid sequences, protein sequences, and antibodies
CC   against them, can be used in the manufacture of a composition. The
CC   composition can be used as a medicament (or in the manufacture of a
CC   medicament) for treating, preventing or diagnosing infection due to
CC   Neisserial bacteria. For example, some of the identified proteins could
CC   be components of vaccines against Meningococcus B; against all serotypes;
CC   and/or against all pathogenic Neissariae. Identification of sequences
CC   from the bacterium will also facilitate production of biological probes,
CC   particularly organism-specific probes. Attempts to make efficacious
CC   Meningococcus B vaccines have failed mainly due to antigen tolerance.
CC   Multivalent vaccines have also been tried but none have successfully
CC   overcome antigenic variability. The provision of further, complete
CC   sequences may provide an opportunity to identify secreted or surface
CC   exposed proteins that may be presumed targets for the immune system and
CC   which are not antigenically variable or at least more conserved than
CC   other more variable regions.
XX
```

```
SQ   Sequence 72750 BP; 17518 A; 19945 C; 18810 G; 16477 T; 0 other;


  Query Match             6.3%;  Score 37.6;  DB 21;  Length 72750;
  Best Local Similarity   55.3%;  Pred. No. 3.2;
  Matches   73;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy     395 cgacgaggtctacggcggtgacgtccacgactaccgccccgcagcgaacaagtccatacg 454
           ||| | ||   |   |  | ||| |  ||||| ||| | |   |  || | ||   | |
Db    9250 cgaagtggatgaactcactcacggcggcgacttccggcacttcggcaaaacgttttttca 9309

Qy     455 ggcacaatggccgacctgtgacggctggattggtggcagctaatagtggtgccagcgcgg 514
           | ||  ||       || ||||  |  |||||||||  |||   |  | |   |  ||||
Db    9310 gccaatattcgatggctttgacatcgtgattggtggtggcttcgatggctttgacggcgg 9369

Qy     515 cctcgtctccca 526
           || ||||| |||
Db    9370 ccgcgtcttcca 9381


RESULT  13
AAF21544/c
ID   AAF21544 standard; DNA; 349980 BP.
XX
AC   AAF21544;
XX
DT   13-MAR-2001  (first entry)
XX
DE   Neisseria meningitidis B nucleotide sequence SEQ ID NO:1.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; immunogenic; vaccine;
KW   diagnosis; antigen; detection; infection; gene therapy; antibacterial;
KW   ds.
XX
OS   Neisseria meningitidis.
XX
PN   WO200066791-A1.
XX
PD   09-NOV-2000.
XX
PF   08-MAR-2000; 2000WO-US05928.
XX
PR   30-APR-1999;   99US-0132068.
PR   08-OCT-1999;   99WO-US23573.
PR   28-FEB-2000; 2000GB-0004695.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Pizza M,  Hickey E,  Peterson J,  Tettelin H,  Venter JC,  Masignani V;
PI   Galeotti C,  Mora M,  Ratti G,  Scarselli M,  Scarlato V,  Rappuoli R;
PI   Frazer CM,  Grandi G;
XX
DR   WPI; 2000-647603/62.
XX
PT   Neisseria meningitidis B full length genome sequence and open reading
PT   frames are used to detect, treat and prevent Neisserial infections -
XX
PS   Claim 7; Appendix A; 692pp; English.
XX
CC   The present invention describes the full length genome of
CC   Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
CC   to AAF21613 represent fragments of the NMB genomic sequence, as the
CC   sequence was too long to go in a record on its own it was split into 8
CC   sequences which overlap each other at the beginning and end of each
CC   sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
CC   the beginning  of AAF21607, the last 49980 bp of AAF21607 are repeated at
CC   the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
CC   Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
CC   AAF21606 represent PCR primers which are used in the exemplification of
CC   the present invention. The NMB genome and fragments from it have
CC   antibacterial activity, and can be used in vaccines and gene therapy.
CC   Neisseria nucleic acids, proteins and/or antibodies which binds to the
CC   proteins can be used in compositions for treating or preventing infection
CC   due to Neisserial bacteria or as a diagnostic reagent for detecting the
CC   presence of Neisserial bacteria or of antibodies raised to Neisserial
CC   bacteria. Computers, computer memory, computer storage medium or computer
CC   databases can be used in a search to identify open reading frames (ORFs)
CC   or coding sequences within the NMB genome. The DNA sequences provide
CC   further opportunities to find antigenic or immunogenic proteins which are
CC   more effective in vaccines than the outer membrane proteins currently
CC   used.
XX
SQ   Sequence 349980 BP; 83241 A; 85091 C; 95206 G; 86442 T; 0 other;


  Query Match             6.3%;  Score 37.6;  DB 21;  Length 349980;
  Best Local Similarity   55.3%;  Pred. No. 4.9;
  Matches   73;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy     395 cgacgaggtctacggcggtgacgtccacgactaccgccccgcagcgaacaagtccatacg 454
           ||| | ||   |   |  | ||| |  ||||| ||| | |   |  || | ||   | |
Db  291610 CGAAGTGGATGAACTCACTCACGGCGGCGACTTCCGGCACTTCGGCAAAACGTTTTTTCA 291551

Qy     455 ggcacaatggccgacctgtgacggctggattggtggcagctaatagtggtgccagcgcgg 514
           | ||  ||       || ||||  |  |||||||||  |||   |  | |   |  ||||
Db  291550 GCCAATATTCGATGGCTTTGACATCGTGATTGGTGGTGGCTTCGATGGCTTTGACGGCGG 291491

Qy     515 cctcgtctccca 526
           || ||||| |||
Db  291490 CCGCGTCTTCCA 291479


RESULT  14
AAX33558
ID   AAX33558 standard; DNA; 921 BP.
XX
AC   AAX33558;
XX
DT   07-JUL-1999  (first entry)
XX
DE   Rice beta-glucanase Gns6 encoding DNA.
XX
KW   Rice; beta-glucanase; Gns; promoter; plant resistance; expression;
KW   fungal infection; transgenic monocotyledon; growth; ss.
XX
OS   Oryza sativa.
XX
PN   WO9859046-A1.
XX
PD   30-DEC-1998.
XX
PF   25-JUN-1998;   98WO-US13525.
XX
PR   25-JUN-1997;   97US-0050675.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Rodriguez RL;
XX
DR   WPI; 1999-105620/09.
DR   P-PSDB; AAY09295.
XX
PT   New polynucleotides which hybridise with rice beta-glucanase genes -
PT   useful for transforming monocot plants for various characteristics
PT   including increased resistance to fungal infection
XX
PS   Claim 4; Page 51-52; 90pp; English.
XX
CC   The present invention describes isolated polynucleotides (PN) comprising
CC   a sequence which hybridises under high stringency with a rice
CC   beta-glucanase gene. Also described are: (A) a chimeric gene (CG1) used
CC   to produce a transgenic monocot plant, comprising: (i) a transcriptional
CC   regulatory region which hybridises under high stringency with a rice beta
CC   -glucanase promoter; and (ii) a DNA sequence heterologous to the
CC   regulatory region, and encoding a protein to be produced by the plant;
```

CC and (iii) a second DNA sequence encoding a signal polypeptide operatively
CC linked 5' to 3', so that the signal polypeptide is in translation frame
CC with the protein, and is effective to facilitate secretion of the protein
CC across aleurone or scutellar epithelium layers into the endosperm of
CC seeds obtained from the plant; (B) a monocot plant stably transformed
CC with CG1; (C) seeds from the above plant; (D) a method of enhancing the
CC resistance of a monocot plant to fungal infection by stably transforming
CC it with CG1; (E) a method of producing a heterologous protein, comprising
CC stably transforming a plant with CG1, obtaining seeds from the
CC transformed plant, germinating the seeds and obtaining the protein from
CC the seed endosperm; (F) an isolated protein having the characteristics of
CC a rice beta-glucanase enzyme, corresponding to beta-glucanases 2-9
CC (Gns 2-9). The polynucleotides of the invention are useful to transform
CC monocot plants. The polynucleotides and the proteins produced may also
CC provide increased resistance to fungal infection, improved growth
CC characteristics, and high expression levels of heterologous proteins in
CC various tissues obtained from the plant.
XX
SQ Sequence 921 BP; 181 A; 308 C; 287 G; 145 T; 0 other;

Query Match 6.3%; Score 37.2; DB 20; Length 921;
Best Local Similarity 52.6%; Pred. No. 1.3;
Matches 81; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

Qy 378 cactggagcgccggccacgacgaggtctacggcggtgacgtccacgactaccgccccgca 437
Db 256 cgctacatcgccgtcggcaacgaggtcaccggcgacgacacgggcaacatcctcccggcc 315

Qy 438 gcgaacaagtccatacgggcacaatggccgacctgtgacggctggattggtggcagctaa 497
Db 316 atgaagaacctcaacgccgcgctcggcgcggccggcctcggcggcgtcggggtgtcgacg 375

Qy 498 tagtggtgccagcgcggcctcgtctcccacagcc 531
Db 376 tcggtgtcccagggcgtgatcgccaactcctacc 409

RESULT 15
AAX33568
ID AAX33568 standard; DNA; 996 BP.
XX
AC AAX33568;
XX
DT 07-JUL-1999 (first entry)
XX
DE Rice Gns6 CDS.
XX
KW Rice; beta-glucanase; Gns; promoter; plant resistance; expression;
KW fungal infection; transgenic monocotyledon; growth; ss.
XX
OS Oryza sativa.
XX
PN WO9859046-A1.
XX
PD 30-DEC-1998.
XX
PF 25-JUN-1998; 98WO-US13525.
XX
PR 25-JUN-1997; 97US-0050675.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Rodriguez RL;
XX
DR WPI; 1999-105620/09.
DR P-PSDB; AAY09303.
XX
PT New polynucleotides which hybridise with rice beta-glucanase genes -
PT useful for transforming monocot plants for various characteristics
PT including increased resistance to fungal infection
XX
PS Disclosure; Page 69-71; 90pp; English.

XX
CC The present invention describes isolated polynucleotides (PN) comprising
CC a sequence which hybridises under high stringency with a rice
CC beta-glucanase gene. Also described are: (A) a chimeric gene (CG1) used
CC to produce a transgenic monocot plant, comprising: (i) a transcriptional
CC regulatory region which hybridises under high stringency with a rice beta
CC -glucanase promoter; and (ii) a DNA sequence heterologous to the
CC regulatory region, and encoding a protein to be produced by the plant;
CC and (iii) a second DNA sequence encoding a signal polypeptide operatively
CC linked 5' to 3', so that the signal polypeptide is in translation frame
CC with the protein, and is effective to facilitate secretion of the protein
CC across aleurone or scutellar epithelium layers into the endosperm of
CC seeds obtained from the plant; (B) a monocot plant stably transformed
CC with CG1; (C) seeds from the above plant; (D) a method of enhancing the
CC resistance of a monocot plant to fungal infection by stably transforming
CC it with CG1; (E) a method of producing a heterologous protein, comprising
CC stably transforming a plant with CG1, obtaining seeds from the
CC transformed plant, germinating the seeds and obtaining the protein from
CC the seed endosperm; (F) an isolated protein having the characteristics of
CC a rice beta-glucanase enzyme, corresponding to beta-glucanases 2-9
CC (Gns 2-9). The polynucleotides of the invention are useful to transform
CC monocot plants. The polynucleotides and the proteins produced may also
CC provide increased resistance to fungal infection, improved growth
CC characteristics, and high expression levels of heterologous proteins in
CC various tissues obtained from the plant.
XX
SQ Sequence 996 BP; 194 A; 330 C; 305 G; 167 T; 0 other;

Query Match 6.3%; Score 37.2; DB 20; Length 996;
Best Local Similarity 52.6%; Pred. No. 1.3;
Matches 81; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

Qy 378 cactggagcgccggccacgacgaggtctacggcggtgacgtccacgactaccgccccgca 437
Db 331 cgctacatcgccgtcggcaacgaggtcaccggcgacgacacgggcaacatcctcccggcc 390

Qy 438 gcgaacaagtccatacgggcacaatggccgacctgtgacggctggattggtggcagctaa 497
Db 391 atgaagaacctcaacgccgcgctcggcgcggccggcctcggcggcgtcggggtgtcgacg 450

Qy 498 tagtggtgccagcgcggcctcgtctcccacagcc 531
Db 451 tcggtgtcccagggcgtgatcgccaactcctacc 484

Search completed: September 21, 2001, 02:41:41
Job time: 11397 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: September 20, 2001, 16:22:20 ; Search time 4309.39 Seconds
(without alignments)
1302.967 Million cell updates/sec

Title: US-09-138-735-1_COPY_1232_1825
Perfect score: 594
Sequence: 1 caggtacagcgtaacggctt.....aaacgagctccgtggcgata 594

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est75:*
137: gb_est76:*
138: gb_est77:*
139: gb_est78:*
140: gb_est79:*
141: gb_est80:*
142: gb_est81:*
143: gb_est82:*
144: gb_est83:*
145: gb_est84:*
146: gb_est85:*
147: gb_est86:*
148: gb_est87:*
149: gb_est88:*
150: gb_est89:*
151: gb_est90:*
152: gb_est99:*
153: gb_est100:*
154: gb_est101:*
155: gb_est102:*
156: gb_est103:*
157: gb_est104:*
158: gb_est105:*
159: gb_est106:*
160: gb_est67:*
161: gb_est68:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_esthum29:*
177: em_esthum30:*
178: em_esthum31:*
179: em_esthum32:*
180: em_esthum33:*
181: em_estom3:*
182: em_estpl11:*
183: em_estro21:*
184: em_estro22:*
185: em_estro23:*
186: em_htc:*
187: gb_est107:*
188: gb_est108:*
189: gb_est109:*
190: gb_est110:*
191: gb_est111:*
192: gb_htc:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_other:*
207: em_gss_pln1:*
208: em_gss_pln2:*
209: em_gss_pro:*
210: em_gss_rod1:*
211: em_gss_rod2:*
212: em_gss_rod3:*
213: em_gss_rod4:*
214: em_gss_rod5:*
215: em_gss_vrt1:*
216: em_gss_vrt2:*
217: em_gss_vrt3:*
218: gb_gss1:*
219: gb_gss2:*
220: gb_gss3:*
221: gb_gss4:*
222: gb_gss5:*
223: gb_gss6:*
224: gb_gss7:*
225: gb_gss8:*
226: gb_gss9:*
227: gb_gss10:*
228: gb_gss11:*
229: gb_gss12:*
230: gb_gss13:*
231: gb_gss14:*
232: gb_gss15:*
233: gb_gss16:*
234: gb_gss17:*
235: gb_gss18:*
236: gb_gss19:*
237: gb_gss20:*
238: gb_gss21:*
239: gb_gss22:*
240: gb_gss23:*
241: gb_gss24:*
242: gb_gss25:*
243: gb_gss26:*
244: gb_gss27:*
245: gb_gss28:*
246: gb_gss29:*
247: gb_gss30:*
248: gb_gss31:*
249: gb_gss32:*
250: gb_gss33:*
251: gb_gss34:*
252: em_gss_inv4:*
253: em_gss_rod6:*
254: em_gss_rod7:*
255: em_gss_rod8:*
256: gb_gss35:*
257: gb_gss36:*
258: gb_gss37:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	47.8	8.0	925	219	CNS0091P	AL053013 Drosophil
2	44	7.4	466	122	AW922844	AW922844 DG1_46_C0
3	43.8	7.4	1036	219	CNS010BS	AL098770 Drosophil
4	43.4	7.3	558	119	AW677864	AW677864 WS1_11_G0
5	43.4	7.3	1065	219	CNS0108S	AL098662 Drosophil
c 6	42.2	7.1	390	145	BF152676	BF152676 uz94e04.y
7	41.8	7.0	391	137	BE595680	BE595680 PI1_55_C0
8	41.8	7.0	488	120	AW747489	AW747489 WS1_68_D0
9	41.8	7.0	554	119	AW677991	AW677991 WS1_12_H0
10	41.6	7.0	500	30	AV437560	AV437560 AV437560
11	41.4	7.0	487	154	BG466260	BG466260 EtESTed50
12	41.4	7.0	559	247	AZ661219	AZ661219 1M0539M10
c 13	41.2	6.9	447	12	AA858891	AA858891 UI-R-A0-b
14	40.4	6.8	925	219	CNS0091P	AL053013 Drosophil
15	40.2	6.8	332	166	BE363235	BE363235 WS1_61_G0
16	40.2	6.8	339	122	AW925156	AW925156 WS1_76_B0
17	40.2	6.8	352	137	BE595557	BE595557 PI1_53_B0
18	40.2	6.8	353	166	BE363396	BE363396 WS1_62_E0
19	40.2	6.8	410	137	BE592380	BE592380 WS1_94_A0
20	40.2	6.8	412	166	BE356035	BE356035 DG1_121_D
21	40.2	6.8	418	166	BE363254	BE363254 WS1_61_F0
22	40.2	6.8	444	166	BE362725	BE362725 DG1_88_B0
23	40.2	6.8	447	122	AW922332	AW922332 DG1_17_C0
24	40.2	6.8	448	155	BG557134	BG557134 EM1_43_A0
25	40.2	6.8	459	120	AW745877	AW745877 WS1_38_A0
26	40.2	6.8	460	166	BE363431	BE363431 WS1_62_H0
27	40.2	6.8	462	137	BE592941	BE592941 WS1_92_C0
28	40.2	6.8	480	137	BE592157	BE592157 WS1_89_B0
29	40.2	6.8	489	137	BE594042	BE594042 WS1_101_E
30	40.2	6.8	489	166	BE363359	BE363359 WS1_62_A0
31	40.2	6.8	490	154	BG464372	BG464372 EM1_69_A0
32	40.2	6.8	493	152	BG356444	BG356444 EM1_22_F0
33	40.2	6.8	497	119	AW677984	AW677984 WS1_12_H1
34	40.2	6.8	498	155	BG556692	BG556692 EM1_40_D0
35	40.2	6.8	499	166	BE363274	BE363274 WS1_61_E0
36	40.2	6.8	500	141	BE917797	BE917797 OV1_7_C02
37	40.2	6.8	502	137	BE593867	BE593867 WS1_103_D
38	40.2	6.8	502	166	BE356939	BE356939 DG1_145_H
39	40.2	6.8	503	137	BE593309	BE593309 WS1_100_B
40	40.2	6.8	506	137	BE597160	BE597160 PI1_71_E0
41	40.2	6.8	506	155	BG556263	BG556263 EM1_67_G0
42	40.2	6.8	506	166	BE360151	BE360151 DG1_61_D0
43	40.2	6.8	506	166	BE363530	BE363530 WS1_63_A0
44	40.2	6.8	507	137	BE592188	BE592188 WS1_89_E0
45	40.2	6.8	509	119	AW678532	AW678532 WS1_16_C0

ALIGNMENTS

```
RESULT  1
CNS0091P/c
LOCUS       CNS0091P      925 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL053013
VERSION     AL053013.1  GI:4934461
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 925)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .925
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR19D16"
                     /note="end : TET3"
BASE COUNT      120 a     61 c     61 g    172 t    511 others
ORIGIN

  Query Match             8.0%;  Score 47.8;  DB 219;  Length 925;
  Best Local Similarity   11.8%;  Pred. No. 0.024;
  Matches   37;  Conservative  151;  Mismatches  125;  Indels    0;  Gaps    0;

Qy     256 ccagccaaggcagccgcgcctcctgccgcggcgcgatcggctgagccgcacgtggggagc 315
           :: :|:  : : ::: :::| :| ::::::: ::   :::: : |:::: :||::  ::|
Db     914 SSBSCSSSSMSTSSSNSSBCSSCSSSBSSSSTSSMSSSSBSSSSGSSSSSSGTSSACVKC 855

Qy     316 aagatcattgctaatctagtgaatcagctggggattaatgtcacccaaaggagcgtcgtc 375
           |:  |   || :|:| :  :   | |: : :|:: :| |         |||: :::: :
Db     854 NASSSCGCCGCGMABCCMCSSSSSCCGSASARGVKVRASGGAGKRGGGSGGASASHSSSS 795

Qy     376 agcactggagcgccggccacgacgaggtctacggcggtgacgtccacgactaccgccccg 435
           | : : :: : :| : :: ::|:: :: :    | :| : :: :: ::   | ::::: :
Db     794 ACBSSSSSCSASCWSASSSSSASSSRSRSGGGAGGSGASSSRSSSSSSSASAGSVVSSAS 735

Qy     436 cagcgaacaagtccatacgggcacaatggccgacctgtgacggctggattggtggcagct 495
           : :::  : :: :::  : : : :   : : : : :: : :|:  : : :  |:  :| :
Db     734 SSSSSCSSSVSCSSVASSMSCSSBSSSSASASSSSSSSSASCASCSCCCTSWSCSCSTSA 675

Qy     496 aatagtggtgccagcgcggcctcgtctcccacagccgcggcgaaaccaacaggagaagaa 555
           : |  :: ::: :|:::::: ::   :::|: :::: ::::    |  :::: |  | :
Db     674 SMSAARSSSSSSSSCSSSSMSASSSASSSSASSSSSSGSSSSSGSACGBSMSSGGGSGSV 615

Qy     556 aaggcctccgcgg 568
           |:: : ::::::
Db     614 SASSGMSSSVSSS 602


RESULT  2
AW922844
LOCUS       AW922844      466 bp    mRNA            EST       19-JUL-2000
DEFINITION  DG1_46_C04.g1_A002 Dark Grown 1 (DG1) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   AW922844
VERSION     AW922844.1  GI:8088669
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 466)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
```

```
            ,L.H.
  TITLE     An EST database from Sorghum: dark-grown seedlings
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: PolyTMix
            High quality sequence start: 135
            High quality sequence stop: 466
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .466
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Dark Grown 1 (DG1)"
                     /note="Organ: 5-day-old dark-grown seedlings; Vector:
                     Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was
                     made from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT       67 a    121 c    182 g     95 t      1 others
ORIGIN
```

```
  Query Match             7.4%;  Score 44;  DB 122;  Length 466;
  Best Local Similarity   49.8%;  Pred. No. 0.22;
  Matches  110;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy    210 gacttctgccgctgctgcatcccctgcatcacccctgtttcagcgccagccaaggcagc 269
          | || | |  || |  ||  |  || |  | ||||| ||    | | | || | ||  |
Db     30 GGCTACGGAGGCGGTGGCGGCGGCTACGGCGCCCCCCGTGAGGGTGNCGGCTACGGAGGA 89

Qy    270 cgcgcctcctgccgcggcgcgatcggctgagccgcacgtggggagcaagatcattgctaa 329
          ||| |  | ||  ||||| |  | ||   |  | | | ||  |   ||    | ||||
Db     90 GGCGGCGGCGGCTACGGCGGGCGCCGCGAGGGAGGAGGCGGCTACGGAGGCGGTGGCTAC 149

Qy    330 tctagtgaatcagctggggattaatgtcacccaaaggagcgtcgtcagcactggagcgcc 389
          | | |    |  ||  |    ||| ||   |||   |  |     |  |  ||| |
Db    150 GGTGGCGGCGGCGGCGGCTACGGCGGTCCCCGTGAGGGCGGCGGCGGCTATGGCGGCGGC 209

Qy    390 ggccacgacgaggtctacggcggtgacgtccacgactaccg 430
          |||  | |||  | | || ||||||||  |  || | || |
Db    210 GGCGGCTACGGTGGCAACCGCGGTGACTCCGGCGGCAACTG 250
```

```
RESULT   3
CNS010BS
LOCUS       CNS010BS     1036 bp    DNA             GSS       26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN03H11 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL098770
VERSION     AL098770.1  GI:5610381
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Plasmid Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1036)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -: This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1036
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN03H11"
                     /note="end : SP6"
BASE COUNT      240 a    127 c    155 g    234 t    280 others
ORIGIN
```

```
  Query Match             7.4%;  Score 43.8;  DB 219;  Length 1036;
  Best Local Similarity   23.6%;  Pred. No. 0.27;
  Matches   92;  Conservative  113;  Mismatches  185;  Indels    0;  Gaps    0;

Qy     94 cggtcctcccgcagggaagaacagcagccaggccaaaaaacatcggtagtggcgacggcg 153
          |:  || :: ::::|  |:|| |  ::  : |: ||||||  | | | |: :::: : :||
Db    637 CKACCCSSSAVSMSGASARAAAAAAVSAAVAGMGAAAAAAGASCYGGASGRRSSSSCRCG 696

Qy    154 aaaccggggtgtgtgtcctcgggcactgacgcggcgagtagcagtcataccaatacgact 213
            :|:  :       :    |      || :  :   : ||||:    |:: |  : |:
Db    697 CSVCSYCKCGCGCCCYGAAAGAAAGACGAGSAAVGCGSAAGCASSAGCAMMGASCSAASA 756

Qy    214 tctgccgctgctgcatcccctgcatcacccctgtttcagcgccagccaaggcagccgcg 273
              |  :| :|: | :|     | : : : |   : :|:::  : ||| :|: :|
Db    757 RAVCGCCGCSCGSCVSCASCSCGCSCGSGMGSCGAMASSSCSSSCASWAAGASASGSGGC 816

Qy    274 cctcctgccgcggcgcgatcggctgagccgcacgtggggagcaagatcattgctaatcta 333
          :| |  ::|:::|:       :: ||||: | :  :| | :: :|  :|  |:  |
Db    817 SCSCGASSCVSSGSCGCGMGCVSCGAGCSCCSSCGSGCGSSSGMGCSSASGGSGSASASS 876

Qy    334 gtgaatcagctggggattaatgtcacccaaaggagcgtcgtcagcactggagcgccggcc 393
          : :     :: ::|    :: : : ::    ::|  |  :  |||:  |: ||||   :|
Db    877 SGSTSCGSSSGSSGCGASRMASGSGSSGGSGSVAARGCGSARAGCRRAGSSGCGCACNSC 936

Qy    394 acgacgaggtctacggcggtgacgtccacgactaccgccccgcagcgaacaagtccatac 453
          |:  |: |    |::::|   |     | :|:   :::|: :  ::|: :    : : |
Db    937 ASMGCKCGMAAAASVSSGMGCAAAAAAAARASGCGSSSCSGSGGSVGVSBSTCGSARVAR 996

Qy    454 gggcacaatggccgacctgtgacggctgga 483
          |:|  :|  |:::| :: : : :|:: |||
Db    997 GSGGCSAGCGSSSGGSSVSSVGSGSSGGGA 1026
```

```
RESULT   4
AW677864
LOCUS       AW677864      558 bp    mRNA            EST       19-JUL-2000
DEFINITION  WS1_11_G05.g1_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA,
            mRNA sequence.
ACCESSION   AW677864
VERSION     AW677864.1  GI:7551577
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 558)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: water-stressed plants
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
```

```
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: T7
            High quality sequence start: 146
            High quality sequence stop: 558
            POLYA=Yes.
FEATURES              Location/Qualifiers
     source           1. .558
                      /organism="Sorghum bicolor"
                      /db_xref="taxon:4558"
                      /clone_lib="Water-stressed 1 (WS1)"
                      /note="Organ: Mix of 5-week old plants on days 7 & 8 after
                      water was withheld; Vector: Lambda Zap; Site_1: XhoI;
                      Site_2: EcoRI; The library was made from poly-A RNA in the
                      cloning vector lambda ZAP II.  Clones to be sequenced were
                      prepared by mass excision."
BASE COUNT       88 a    175 c    210 g     85 t
ORIGIN


  Query Match             7.3%;  Score 43.4;  DB 119;  Length 558;
  Best Local Similarity   49.8%;  Pred. No. 0.33;
  Matches  110;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy     210 gacttctgccgctgctgcatcccctgcatcacccctgtttcagcgccagccaaggcagc 269
           | || | |  || |  ||  |  || |  | ||||| ||    | | | || | ||  |
Db     255 GGCTACGGAGGCGGTGGCGGCGGCTACGGCGCCCCCCGTGAGGGTGGCGGCTACGGAGGA 314

Qy     270 cgcgcctcctgccgcggcgcgatcggctgagccgcacgtggggagcaagatcattgctaa 329
           |||  |  | ||  ||||| |  | ||   |  | |   ||  |  |||    | ||||
Db     315 GGCGGCGGCGGCTACGGCGGGCGCCGCGAGGGAGGAGCCGGCTACGAAGGCGGTGGCTAC 374

Qy     330 tctagtgaatcagctggggattaatgtcacccaaaggagcgtcgtcagcactggagcgcc 389
           | | |     |  ||  |     ||| ||   |||   |  |     |  |  ||| |
Db     375 GGTGGCGGCGGCGGCGGCTACGGCGGTCGCCGTGAGGGCGGCGGCGGCTATGGCGGCGGC 434

Qy     390 ggccacgacgaggtctacggcggtgacgtccacgactaccg 430
           |||  | |||  | | || ||||||||  |  || | || |
Db     435 GGCGGCTACGGTGGCAACCGCGGTGACTCCGGCGGCAACTG 475


RESULT   5
CNS0108S
LOCUS       CNS0108S     1065 bp    DNA             GSS       26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN03J03 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL098662
VERSION     AL098662.1  GI:5610273
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Plasmid Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1065)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES              Location/Qualifiers
     source           1. .1065
                      /organism="Drosophila melanogaster"
                      /plasmid="pBeloBAC11"
                      /db_xref="taxon:7227"
                      /clone_lib="DrosBAC"
                      /clone="BACN03J03"
                      /note="end : T7"
BASE COUNT      121 a    225 c    252 g    313 t    154 others
ORIGIN


  Query Match             7.3%;  Score 43.4;  DB 219;  Length 1065;
  Best Local Similarity   27.4%;  Pred. No. 0.35;
  Matches   51;  Conservative   62;  Mismatches   73;  Indels    0;  Gaps    0;

Qy     141 agtggcgacggcgaaaccggggtgtgtgtcctcgggcactgacgcggcgagtagcagtca 200
           :: :|:: ::::|   :  :  |:||:|    ::| |  :| ::::: : :: |: |:
Db     755 RVGSGSSGSSSSSGSGSSGCSTCTSTGKGGGGSSSGCCCGKGSSSSKBGSSBKSGSGGKGT 814

Qy     201 taccaatacgacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgccagc 260
           |||  |  |  |:| :   :    || :|||| |  | ||||      |: | || :|
Db     815 CACCTTTBKGTKTWCGKTTCSCCGGGCYYCCCCCCCCCCTCCCCGSCGNCCMCCCCCCSC 874

Qy     261 caaggcagccgcgcctcctgccgcggcgcgatcggctgagccgcacgtggggagcaagat 320
           |: | |  :::: |: |: ::::::::::  ||: | : |::|: :| |::: ||  |::
Db     875 CMTGCCYTSSSSCCSSCSSSSSSSSSSSCCTCKCCSKTGSSGSSSGGGSSSCGCTKGVK 934

Qy     321 cattgc 326
           : |:
Db     935 ACKGGS 940


RESULT   6
BF152676/c
LOCUS       BF152676      390 bp    mRNA            EST       26-OCT-2000
DEFINITION  uz94e04.y1 NCI_CGAP_Te1 Mus musculus cDNA clone IMAGE:3809358 5',
            mRNA sequence.
ACCESSION   BF152676
VERSION     BF152676.1  GI:11034071
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 390)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov

            MGI:1451470
            Seq primer: -40RP from Gibco
            High quality sequence stop: 371.
FEATURES              Location/Qualifiers
     source           1. .390
                      /organism="Mus musculus"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:3809358"
                      /clone_lib="NCI_CGAP_Te1"
                      /sex="male"
                      /dev_stage="5 months"
                      /lab_host="DH10B (T1 phage-resistant)"
                      /note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
```

Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.3 kb. Library constructed by Life Technologies."
BASE COUNT 49 a 150 c 135 g 56 t
ORIGIN

Query Match 7.1%; Score 42.2; DB 145; Length 390;
Best Local Similarity 54.1%; Pred. No. 0.65;
Matches 86; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

Qy 137 cggtagtggcgacggcgaaaccggggtgtgtgtcctcgggcactgacgcggcgagtagca 196
Db 172 CGGCAGCGCCCACCGCACCGCCTGGGCCTCTGGACTGGGGCTGCGGCTGCGCGACCCCAA 113

Qy 197 gtcataccaatacgacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgc 256
Db 112 GGCGTCCAGAAGCAGCCTCCGCCGCTGCTGCAGCAGCTCCTGACGCTGCTGTCTCCTCCT 53

Qy 257 cagccaaggcagccgcgcctcctgccgcggcgcgatcgg 295
Db 52 GTGCAGCAGCTGCTGCCACAGCTCCCGCCCCGTGCCCGG 14

RESULT 7
BE595680
LOCUS BE595680 391 bp mRNA EST 18-AUG-2000
DEFINITION PI1_55_C06.g1_A002 Pathogen induced 1 (PI1) Sorghum bicolor cDNA, mRNA sequence.
ACCESSION BE595680
VERSION BE595680.1 GI:9850753
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 391)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Dean,R., Sudman,M. and Pratt ,L.H.
TITLE An EST database from Sorghum: pathogen-induced plants
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
Seq primer: PolyTMix
High quality sequence start: 2
High quality sequence stop: 260
POLYA=No.
FEATURES Location/Qualifiers
source 1. .391
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Pathogen induced 1 (PI1)"
/note="Organ: Anthracnose-infected leaves from two-week-old sorghum plants 48 hr after inoculation; Vector: pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2: EcoRI; Two-week-old sorghum plants (BTX 623 cultivar) were infected with pathogen (isolate FRM42I of Colletotrichum graminicola, which is a sorghum isolate). RNA was prepared from infected leaves harvested from 45 seedlings 48 hours after inoculation. Note: young seedlings (2 weeks old) exhibit juvenile resistant reaction, which is an incompatible interaction. As they grow older (4 weeks or older), plants resume susceptiblity to anthracnose disease. The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision. WARNING: While most or all ESTs are expected to derive from the host plant, no effort was made to eliminate ESTs deriving from the pathogen."
BASE COUNT 53 a 99 c 161 g 78 t
ORIGIN

Query Match 7.0%; Score 41.8; DB 137; Length 391;
Best Local Similarity 49.3%; Pred. No. 0.83;
Matches 109; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

Qy 210 gacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaaggcagc 269
Db 12 GGCTACGGAGGCGGTGGCGGCGGCTACGGCGGCCGCCGTGAGGGTGGCGGCTACGGAGGA 71

Qy 270 cgcgcctcctgccgcggcgcgatcggctgagccgcacgtggggagcaagatcattgctaa 329
Db 72 GGCGGCGGCGGCTACGGCGGGCTCCGCGAGGGAGGAGGCGGCTACGGAGGCGGTGGCTAC 131

Qy 330 tctagtgaatcagctggggattaatgtcacccaaaggagcgtcgtcagcactggagcgcc 389
Db 132 GGTGGCGGCGGCGGCGGCTACGGCGGTCGCCGTGAGGGCGGCGGCGGCTATGGCGGCGGC 191

Qy 390 ggccacgacgaggtctacggcggtgacgtccacgactaccg 430
Db 192 GGCGGCTACGGTGGCAACCGCGGTGACTCCGGCGGCAACTG 232

RESULT 8
AW747489
LOCUS AW747489 488 bp mRNA EST 19-JUL-2000
DEFINITION WS1_68_D09.g1_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA, mRNA sequence.
ACCESSION AW747489
VERSION AW747489.1 GI:7661227
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 488)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt ,L.H.
TITLE An EST database from Sorghum: water-stressed plants
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
Seq primer: T7
High quality sequence start: 4
High quality sequence stop: 488
POLYA=Yes.
FEATURES Location/Qualifiers
source 1. .488
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Water-stressed 1 (WS1)"
/note="Organ: Mix of 5-week old plants on days 7 & 8 after water was withheld; Vector: Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision."
BASE COUNT 73 a 138 c 189 g 88 t

```
ORIGIN


  Query Match             7.0%;  Score 41.8;  DB 120;  Length 488;
  Best Local Similarity   49.3%;  Pred. No. 0.85;
  Matches  109;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps    0;

Qy     210 gacttctgccgctgctgcatcccctgcatcacccctgtttcagcgccagccaaggcagc 269
           | || | |  || |  ||  |  || | |  |||| ||    | | | || | ||  |
Db     113 GGCTACGGAGGCGGTGGCGGCGGCTACGGCGGCCCCCGTGAGGGTGGCGGCTACGGAGGA 172

Qy     270 cgcgcctcctgccgcggcgcgatcggctgagccgcacgtggggagcaagatcattgctaa 329
           ||| |  | ||  ||||| |  | ||   |  | | | ||  |   ||    | ||||
Db     173 GGCGGCGGCGGCTACGGCGGGCCCCGCGAGGGAGGAGGCGGCTACGGAGGCGGTGGCTAC 232

Qy     330 tctagtgaatcagctggggattaatgtcacccaaaggagcgtcgtcagcactggagcgcc 389
           | | |     |  ||  |     ||| ||   |||   |  |     |  |  ||| |
Db     233 GGTGGCGGCGGCGGCGGCTACGGCGGTCCCCGTGAGGGCGGCGGCGGCTATGGCGGCGGC 292

Qy     390 ggccacgacgaggtctacggcggtgacgtccacgactaccg 430
           |||  | |||  | | || ||||||||  |  || | || |
Db     293 GGCGGCTACGGTGGCAACCGCGGTGACTCCGGCGGCAACTG 333


RESULT  9
AW677991
LOCUS       AW677991      554 bp    mRNA            EST       19-JUL-2000
DEFINITION  WS1_12_H04.g1_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA,
            mRNA sequence.
ACCESSION   AW677991
VERSION     AW677991.1  GI:7551708
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 554)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: water-stressed plants
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: T7
            High quality sequence start: 208
            High quality sequence stop: 554
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .554
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Water-stressed 1 (WS1)"
                     /note="Organ: Mix of 5-week old plants on days 7 & 8 after
                     water was withheld; Vector: Lambda Zap; Site_1: XhoI;
                     Site_2: EcoRI; The library was made from poly-A RNA in the
                     cloning vector lambda ZAP II.  Clones to be sequenced were
                     prepared by mass excision."
BASE COUNT       92 a    147 c    205 g    110 t
ORIGIN


  Query Match             7.0%;  Score 41.8;  DB 119;  Length 554;
  Best Local Similarity   49.3%;  Pred. No. 0.87;
  Matches  109;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps    0;

Qy     210 gacttctgccgctgctgcatcccctgcatcacccctgtttcagcgccagccaaggcagc 269
           | || | |  || |  ||  |  || | |  |||| ||    | | | || | ||  |
Db     107 GGCTACGAAGGCGGTGGCGGCGGCTACGGCGGCCCCCGTGAGGGTGGCGGCTACGGAGGA 166

Qy     270 cgcgcctcctgccgcggcgcgatcggctgagccgcacgtggggagcaagatcattgctaa 329
           ||| |  | ||  ||||| |  | ||   |  | | | ||  |   ||    | ||||
Db     167 GGCGGCGGCGGCTACGGCGGGCGCCGCGAGGGAGAAGGCGGCTACGAAGGCGGTGGCTAC 226

Qy     330 tctagtgaatcagctggggattaatgtcacccaaaggagcgtcgtcagcactggagcgcc 389
           | | |     |  ||  |     ||| ||   |||   |  |     |  |  ||| |
Db     227 GGTGGCGGCGGCGGCGGCTACGGCGGTCCCCGTGAGGGCGGCGGCGGCTATGGCGGCGGC 286

Qy     390 ggccacgacgaggtctacggcggtgacgtccacgactaccg 430
           |||  | |||  | | || ||||||||  |  || | || |
Db     287 GGCGGCTACGGTGGCAACCGCGGTGACTCCGGCGGCAACTG 327


RESULT  10
AV437560
LOCUS       AV437560      500 bp    mRNA            EST       23-AUG-2000
DEFINITION  AV437560 Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone
            PS023b06_r 5', mRNA sequence.
ACCESSION   AV437560
VERSION     AV437560.1  GI:8592785
KEYWORDS    EST.
SOURCE      Porphyra yezoensis.
  ORGANISM  Porphyra yezoensis
            Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;
            Porphyra.
REFERENCE   1  (bases 1 to 500)
  AUTHORS   Nikaido,I., Asamizu,E., Nakajima,M., Nakamura,Y., Saga,N. and
            Tabata,S.
  TITLE     Generation of 10,154 expressed sequence tags from a leafy
            gametophyte of a marine red alga, Porphyra yezoensis
  JOURNAL   DNA Res. 7, 223-227 (2000)
  MEDLINE   20363100
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .500
                     /organism="Porphyra yezoensis"
                     /strain="TU-1"
                     /db_xref="taxon:2788"
                     /clone="PS023b06_r"
                     /clone_lib="Porphyra yezoensis TU-1"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT       83 a    190 c    159 g     68 t
ORIGIN


  Query Match             7.0%;  Score 41.6;  DB 30;  Length 500;
  Best Local Similarity   45.2%;  Pred. No. 0.97;
  Matches  152;  Conservative    0;  Mismatches  184;  Indels    0;  Gaps    0;

Qy     253 gcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggctgagccgcacgtgggg 312
           || || |||    |  |  || |  |||| || || |   ||  ||||    ||| |  |
Db     116 GCCCCCGCCGCACCTACGACGGCCGCTGCGGCCGCCCCTACGAATGAGGACGACGAGACG 175

Qy     313 agcaagatcattgctaatctagtgaatcagctggggattaatgtcacccaaaggagcgtc 372
           | |  | |    ||  |  | | | | |      |       || | |    | |  |
Db     176 AACGCGCTGCGCGCGCAGGTGGCGGAGCTCGCCGCCGCCCTTGCCGCTGCCAAGGCCCAA 235

Qy     373 gtcagcactggagcgccggccacgacgaggtctacggcggtgacgtccacgactaccgcc 432
           |   |    ||| || || |  | |  ||   |||| |   | ||| |  |||
Db     236 CTGGCAAAACCGGCGGCGCCCGCTGCCGGTACTGACGCGGAAAGCACGACGGCGTCCGTG 295

Qy     433 ccgcagcgaacaagtccatacgggcacaatggccgacctgtgacggctggattggtggca 492
           |||| ||   |||  |   |||    ||  ||     | ||| || | |    |  | |
```

```
Db    296 CCGCCGCCGCCAACGCGTGACGCATGCACGGGGAAGTCAGTGGCGTCCGCGCCGCCGCCG 355

Qy    493 gctaatagtggtgccagcgcggcctcgtctcccacagccgcggcgaaaccaacaggagaa 552
          | |   |||  ||  || |||    |||     |  ||||| |   ||  ||
Db    356 CCAACGCGTGACGCAGGCACGGGGAAGTCAGTGGCGTCCGCGCTGCCGCCGCCAACGCGT 415

Qy    553 gaaaaggcctccgcggcatgtgaaacgagctccgtg 588
          ||   | | | | ||  | ||     |||| ||||
Db    416 AAACCAGACGCGGAGGAGTCTGCGTTTAGCTGCGTG 451


RESULT  11
BG466260
LOCUS       BG466260      487 bp    mRNA            EST       20-MAR-2001
DEFINITION  EtESTed50h11.y1 Eimeria tenella S5-2 cDNA Neg Selected Eimeria
            tenella cDNA 5', mRNA sequence.
ACCESSION   BG466260
VERSION     BG466260.1  GI:13395235
KEYWORDS    EST.
SOURCE      Eimeria tenella.
  ORGANISM  Eimeria tenella
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
            Eimeria.
REFERENCE   1  (bases 1 to 487)
  AUTHORS   Liberator,P., Diaz,C., Tang,K., Marra,M., Hillier,L., Kucaba,T.,
            Martin,J., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen
            ,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey
            ,N., Schurk,R., Ritter,E., Kohn,S., Florence,N., Shin,T., Jackson
            ,Y., Cardenas,M., McCann,R., Waterston,R., Wilson,R. and Sibley,D.
  TITLE     WashU-Merck Eimeria tenella project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: David Sibley, Ph.D.
            WashU-Merck Eimeria tenella project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Contact David Sibley (toxoest@borcim.wustl.edu) for further
            information relating to organism, libraries, or clone availability.
            Seq primer: -40RP from Gibco
            High quality sequence stop: 403.
FEATURES             Location/Qualifiers
     source          1. .487
                     /organism="Eimeria tenella"
                     /strain="LS18"
                     /db_xref="taxon:5802"
                     /clone_lib="Eimeria tenella S5-2 cDNA Neg Selected"
                     /dev_stage="Sporozoite stage"
                     /lab_host="SOLR"
                     /note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI
                     ; Sporozoites were obtained from in vitro sporulated and
                     excysted oocysts of E.tenella grown in chickens. cDNA was
                     synthesized from poly mRNA using an oligo-dT primer
                     containing a XhoI site. Following second strand synthesis,
                     EcoRI adapters were ligated to the cDNA and products were
                     size-selected on sephacryl S500. The cDNA were ligated to
                     EcoRI/XhoI prepared lambda ZapII (Stratagene). Clones were
                     converted to phagemids by mass excision using ExAssist
                     helper phage and E.coli SOLR cell (Stratagene). Clones
                     were selected by negative hybridization against a pool of
                     overrepresented ESTs (N>=10,from 1682 previous reads).
                     Insert sizes range from 1.2-2.9Kb. The library may contain
                     a small percentage of host or bacterial contaminents."
BASE COUNT       81 a    143 c    131 g     92 t     40 others
ORIGIN


  Query Match             7.0%;  Score 41.4;  DB 154;  Length 487;
  Best Local Similarity   61.7%;  Pred. No. 1.1;
  Matches   66;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy    192 tagcagtcataccaatacgacttctgccgctgctgcatcccctgcatcacccccctgtttc 251
          || | | |   |  || ||  | |||| || |||||| |  ||||| || ||  ||   |
Db     84 TAACTGCCGCTCACATGCGCGTGCTGCAGCAGCTGCAGCAGCTGCAGCAGCCGTTGCAGC 143

Qy    252 agcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggctg 298
          |||  |||||  ||||| ||  || | || || ||   | | ||||
Db    144 AGCTGCAGCCGCTGCAGCTGCTGCTGCGGCAGCTGCAGCAGCCGCTG 190


RESULT  12
AZ661219
LOCUS       AZ661219      559 bp    DNA             GSS       14-DEC-2000
DEFINITION  1M0539M10R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0539M10 R, DNA sequence.
ACCESSION   AZ661219
VERSION     AZ661219.1  GI:11798365
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 559)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0539  row: M  column: 10
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 559.
FEATURES             Location/Qualifiers
     source          1. .559
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0539M10"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT       17 a    250 c    143 g    149 t
ORIGIN
```

```
  Query Match             7.0%;  Score 41.4;  DB 247;  Length 559;
  Best Local Similarity   58.5%;  Pred. No. 1.1;
  Matches   72;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy     165 tgtgtcctcgggcactgacgcggcgagtagcagtcataccaatacgacttctgccgctgc 224
           ||  ||  | | | |||  || ||   |     || | |   | |  |  |||| ||  |
Db      86 TGCCTCTGCTGCCCCTGCTGCTGCCCCTGCTGCTCCTGCTGCTGCTGCCGCTGCTGCCTC 145

Qy     225 tgcatcccctgcatcacccccctgtttcagcgccagccaaggcagccgcgcctcctgccgc 284
           |||  |||||||  |  |||||| | | ||  | ||||| || ||||   || ||||| |
Db     146 TGCTGCCCCTGCTGCTGCCCCTGCTGCTGCTGCTGCCAATGCTGCCGATGCTGCTGCCCC 205

Qy     285 ggc 287
           |||
Db     206 TGC 208


RESULT  13
AA858891/c
LOCUS       AA858891      447 bp    mRNA            EST       03-JUL-1999
DEFINITION  UI-R-A0-bd-d-01-0-UI.s1 UI-R-A0 Rattus norvegicus cDNA clone
            UI-R-A0-bd-d-01-0-UI 3' similar to gb|AA422729|AA422729 vd29b09.s1
            Knowles Solter mouse 2 cell Mus musculus cDNA clone 793913 5', mRNA
            sequence.
ACCESSION   AA858891
VERSION     AA858891.1  GI:4229084
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 447)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     On Mar 10, 1998 this sequence version replaced gi:2948242.
            Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message. cDNA Library Preparation: M.
            Fatima Bonaldo, Ph.D. Clone distribution: clones will be available
            through Research Genetics This clone is also available through the
            I.M.A.G.E. Consortium at LLNL (info@image.llnl.gov). IMAGE
            ID=1776747 The following repetitive elements were found in this
            cDNA sequence: 15-121, >(CAG)n#Simple_repeat 61-179, >(CGG
            )n#Simple_repeat
            Seq primer: M13 Forward
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .447
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-A0-bd-d-01-0-UI"
                     /clone_lib="UI-R-A0"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; This library
                     consists of a mixture of individually tagged normalized
                     libraries constructed from rat placenta, adult lung, brain
                     , liver, kidney, heart, spleen, ovary, and muscle. The tag
                     is a string of 3-5 nucleotides present between the Not I
                     site and the oligo-dT track which allows identification of
                     the library of origin of a clone within the mixture."
BASE COUNT      125 a    150 c    110 g     62 t
ORIGIN


  Query Match             6.9%;  Score 41.2;  DB 12;  Length 447;
  Best Local Similarity   52.3%;  Pred. No. 1.2;
  Matches   91;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy     137 cggtagtggcgacggcgaaaccggggtgtgtgtcctcgggcactgacgcggcgagtagca 196
           |||  | || | |||||    ||| |   | | |  | |   | || |||||| |
Db     177 CGGCGGCGGAGGCGGCGGTGGCGGCGGCGGGGGCTGCTGCTGCGGAGGCGGCGGGGGTTG 118

Qy     197 gtcataccaatacgacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgc 256
           |  | |   | |  || |||  || |||||  |  ||||  |  |  |||   | |
Db     117 CTGCTGCTGCTGCTGCTGCTGGGGCGGCTGCGGCGGCTGCTGCTGCTGCTGGGGCGGTTG 58

Qy     257 cagccaaggcagccgcgcctcctgccgcggcgcgatcggctgagccgcacgtgg 310
           | ||   ||| || |||    |||| || ||     | ||||  |||  || ||
Db      57 CGGCGGCGGCTGCTGCGGGGGCTGCTGCTGCTGTTGCTGCTGCACCGGGCGCGG 4


RESULT  14
CNS0091P
LOCUS       CNS0091P      925 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL053013
VERSION     AL053013.1  GI:4934461
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 925)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .925
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR19D16"
                     /note="end : TET3"
BASE COUNT      120 a     61 c     61 g    172 t    511 others
ORIGIN


  Query Match             6.8%;  Score 40.4;  DB 219;  Length 925;
  Best Local Similarity   11.4%;  Pred. No. 2.1;
  Matches   42;  Conservative  167;  Mismatches  161;  Indels    0;  Gaps    0;

Qy     160 gggtgtgtgtcctcgggcactgacgcggcgagtagcagtcataccaatacgacttctgcc 219
           ::| |:|:| :: :|::| | : |:|:::: :  :|   |   ::    :: : :| : :
```

```
Db    555 SSGSGYGKGCSSGSGBSCSCCSSCSCSSSSCSCCBCCCCCSCSSYCCSSSBSSSKCSSTS 614

Qy    220 gctgctgcatcccctgcatcaccccctgtttcagcgccagccaaggcagccgcgcctcct 279
          :: :|  : :::| |:|  : :|::: : | : : ::: : :  :|: ::::::: |:
Db    615 BSCSCCCSSKSVCGTSCSSSSSCSSSSSSTSSSSTSSSTSKSSSSGSSSSSSSSYTTSKS 674

Qy    280 gccgcggcgcgatcggctgagccgcacgtggggagcaagatcattgctaatctagtgaat 339
          : : :|: :   :|:|| :::: ::  : : ::: :     |::   ::  | :
Db    675 TSASGSGSWSAGGGSGSTGSTSSSSSSSSTSTSSSSVSSGSKSSTBSSGSBSSSGSSSSS 734

Qy    340 cagctggggattaatgtcacccaaaggagcgtcgtcagcactggagcgccggccacgacg 399
          : :::::   |    : : :::       |  :::: :  : :: : |: :|:: :::
Db    735 STSSBBSCTSTSSSSSSSYSSSTCSCCTCCCSYSYSSSTSSSSSTSWGSTSGSSSSSVGT 794

Qy    400 aggtctacggcggtgacgtccacgactaccgccccgcagcgaacaagtccatacgggcac 459
          ::    :  |:    : |||    :: : : :| |: :::    : :|    ||:::
Db    795 SSSSDSTSTCCSCCCYMCTCCSTYBMBCYTSTSCGGSSSSSGKGGVTKCGCGGCGSSSTN 854

Qy    460 aatggccgacctgtgacggctggattggtggcagctaatagtggtgccagcgcggcctcg 519
          ::| ::  :: : : :::: ::  ::: ::: ::      | :: :::|: :::: : ::
Db    855 GMBGTSSACSSSSSSCSSSSVSSSSKSSASSSSVSSSGSSGVSSNSSSASKSSSSGSVSS 914

Qy    520 tctcccacag 529
          : : : :::
Db    915 GSGSGSGSVS 924
```

```
RESULT  15
BE363235
LOCUS       BE363235      332 bp    mRNA            EST       20-JUL-2000
DEFINITION  WS1_61_G04.g1_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA,
            mRNA sequence.
ACCESSION   BE363235
VERSION     BE363235.1  GI:9304792
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 332)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: water-stressed plants
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: PolyTMix
            High quality sequence start: 7
            High quality sequence stop: 330
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .332
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Water-stressed 1 (WS1)"
                     /note="Organ: Mix of 5-week old plants on days 7 & 8 after
                     water was withheld; Vector: Lambda Zap; Site_1: XhoI;
                     Site_2: EcoRI; The library was made from poly-A RNA in the
                     cloning vector lambda ZAP II.  Clones to be sequenced were
                     prepared by mass excision."
BASE COUNT       48 a     98 c    155 g     31 t
ORIGIN
```

```
  Query Match             6.8%;  Score 40.2;  DB 166;  Length 332;
  Best Local Similarity   48.9%;  Pred. No. 2.2;
  Matches  108;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy    210 gacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaaggcagc 269
          | || | |  || |  ||  |  || | |  || | ||    | | | || | ||  |
Db    109 GGCTACGGAGGCGGTGGCGGCGGCTACGGCGGCCGCCGTGAGGGTGGCGGCTACGGAGGA 168

Qy    270 cgcgcctcctgccgcggcgcgatcggctgagccgcacgtggggagcaagatcattgctaa 329
          ||| |  | ||  ||||| |  | ||   |  | | | ||  |   ||    | ||||
Db    169 GGCGGCGGCGGCTACGGCGGGCGCCGCGAGGGAGGAGGCGGCTACGGAGGCGGTGGCTAC 228

Qy    330 tctagtgaatcagctggggattaatgtcacccaaaggagcgtcgtcagcactggagcgcc 389
          | | |      |  ||  |     ||| ||   |||   |  |     |  |  ||| |
Db    229 GGTGGCGGCGGCGGCGGCTACGGCGGTCGCCGTGAGGGCGGCGGCGGCTATGGCGGCGGC 288

Qy    390 ggccacgacgaggtctacggcggtgacgtccacgactaccg 430
          |||  | |||  | | || |||||||| |  || | || |
Db    289 GGCGGCTACGGTGGCAACCGCGGTGACTCCGGCGGCAACTG 329
```

```
Search completed: September 20, 2001, 23:27:46
Job time: 25526 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: September 20, 2001, 18:26:07 ; Search time 164.23 Seconds
(without alignments)
684.715 Million cell updates/sec

Title: US-09-138-735-1_COPY_1232_1825
Perfect score: 594
Sequence: 1 caggtacagcgtaacggctt.....aaacgagctccgtggcgata 594

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn1_7/ptodata/1/ina/5A_COMB.seq:*
2: /cgn1_7/ptodata/1/ina/5B_COMB.seq:*
3: /cgn1_7/ptodata/1/ina/6A_COMB.seq:*
4: /cgn1_7/ptodata/1/ina/6B_COMB.seq:*
5: /cgn1_7/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn1_7/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	594	100.0	3402	1	US-08-480-917-1	Sequence 1, Appli
	2	39.2	6.6	2790	2	US-08-718-661-1	Sequence 1, Appli
c	3	38.2	6.4	2943	1	US-08-042-747A-7	Sequence 7, Appli
	4	37.6	6.3	2115	2	US-08-474-379C-60	Sequence 60, Appl
	5	37.6	6.3	2115	3	US-09-146-249A-60	Sequence 60, Appl
	6	37.6	6.3	2115	3	US-08-206-188B-60	Sequence 60, Appl
	7	36.8	6.2	3777	3	US-09-121-321-15	Sequence 15, Appl
	8	36.8	6.2	3777	4	US-08-933-803A-15	Sequence 15, Appl
c	9	36.6	6.2	203	4	US-09-043-303-7	Sequence 7, Appli
	10	36.6	6.2	913	1	US-08-217-327-3	Sequence 3, Appli
	11	36.6	6.2	913	1	US-07-885-970A-3	Sequence 3, Appli
	12	36.6	6.2	913	1	US-08-298-687A-3	Sequence 3, Appli
	13	36.6	6.2	913	1	US-08-530-797-2	Sequence 2, Appli
	14	36.6	6.2	913	1	US-08-298-829-3	Sequence 3, Appli
	15	36.6	6.2	913	2	US-08-787-335-2	Sequence 2, Appli
	16	36.6	6.2	1984	1	US-07-885-970A-25	Sequence 25, Appl
	17	36.6	6.2	1985	1	US-08-298-687A-25	Sequence 25, Appl
	18	36.6	6.2	1985	1	US-08-298-829-25	Sequence 25, Appl
c	19	36	6.1	2580	3	US-09-050-863-2	Sequence 2, Appli
	20	36	6.1	5452	2	US-09-130-114-1	Sequence 1, Appli
c	21	36	6.1	9600	4	US-08-910-647-1	Sequence 1, Appli
c	22	36	6.1	10596	1	US-07-884-811-15	Sequence 15, Appl
c	23	36	6.1	10596	1	US-07-885-971-15	Sequence 15, Appl
c	24	36	6.1	10596	1	US-08-087-783A-15	Sequence 15, Appl
c	25	36	6.1	10596	1	US-08-194-088B-15	Sequence 15, Appl
c	26	36	6.1	10596	2	US-08-194-087-15	Sequence 15, Appl
c	27	36	6.1	10596	5	PCT-US93-04648-15	Sequence 15, Appl
c	28	35.2	5.9	3376	1	US-08-320-559-29	Sequence 29, Appl
c	29	35.2	5.9	3376	3	US-08-545-860D-29	Sequence 29, Appl
c	30	35.2	5.9	3376	5	PCT-US94-04496-29	Sequence 29, Appl
c	31	35	5.9	168	1	US-08-469-802B-4	Sequence 4, Appli
c	32	35	5.9	168	2	US-08-267-803B-4	Sequence 4, Appli
c	33	35	5.9	171	1	US-08-469-802B-5	Sequence 5, Appli
c	34	35	5.9	171	2	US-08-267-803B-5	Sequence 5, Appli
c	35	35	5.9	195	1	US-08-469-802B-2	Sequence 2, Appli
c	36	35	5.9	195	2	US-08-267-803B-2	Sequence 2, Appli
c	37	35	5.9	234	1	US-08-469-802B-3	Sequence 3, Appli
c	38	35	5.9	234	2	US-08-267-803B-3	Sequence 3, Appli
c	39	35	5.9	336	2	US-07-814-220-3	Sequence 3, Appli
c	40	35	5.9	336	2	US-07-812-421-3	Sequence 3, Appli
	41	34.8	5.9	1299	1	US-07-661-610C-11	Sequence 11, Appl
c	42	34.8	5.9	2001	1	US-08-674-168-24	Sequence 24, Appl
c	43	34.8	5.9	2001	3	US-08-985-908-10	Sequence 10, Appl
c	44	34.8	5.9	2001	3	US-08-852-730-23	Sequence 23, Appl
	45	34.8	5.9	5437	1	US-07-661-610C-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-480-917-1
; Sequence 1, Application US/08480917
; Patent No. 5820864
; GENERAL INFORMATION:
; APPLICANT: PARANHOS-BACCALA, Glaucia
; APPLICANT: LESENECHAL, Mylene
; APPLICANT: JOLIVET, Michel
; TITLE OF INVENTION: NEW TRYPANOSOMA CRUZI ANTIGEN, AND GENE
; TITLE OF INVENTION: ENCODING THE LATTER; THEIR APPLICATION TO THE DETECTION OF
; TITLE OF INVENTION: CHAGAS DISEASE
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oliff & Berridge
; STREET: 700 South Washington Street, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,917
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 36400
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3402 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-480-917-1

Query Match 100.0%; Score 594; DB 1; Length 3402;
Best Local Similarity 100.0%; Pred. No. 1.1e-152;
Matches 594; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caggtacagcgtaacggcttttgcttcaatcgtacagccgacggtagctgcgtcctggct 60

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1232 CAGGTACAGCGTAACGGCTTTTGCTTCAATCGTACAGCCGACGGTAGCTGCGTCCTGGCT 1291

Qy     61 gacatgtcgattcgattgacgatcttccatctccggtcctcccgcagggaagaacagcag 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1292 GACATGTCGATTCGATTGACGATCTTCCATCTCCGGTCCTCCCGCAGGGAAGAACAGCAG 1351

Qy    121 ccaggccaaaaaaacatcggtagtggcgacggcgaaaccggggtgtgtgtcctcgggcact 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1352 CCAGGCCAAAAAACATCGGTAGTGGCGACGGCGAAACCGGGGTGTGTGTCCTCGGGCACT 1411

Qy    181 gacgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatcccctgcatca 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1412 GACGCGGCGAGTAGCAGTCATACCAATACGACTTCTGCCGCTGCTGCATCCCCTGCATCA 1471

Qy    241 ccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggctgag 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1472 CCCCCTGTTTCAGCGCCAGCCAAGGCAGCCGCGCCTCCTGCCGCGGCGCGATCGGCTGAG 1531

Qy    301 ccgcacgtggggagcaagatcattgctaatctagtgaatcagctggggattaatgtcacc 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1532 CCGCACGTGGGGAGCAAGATCATTGCTAATCTAGTGAATCAGCTGGGGATTAATGTCACC 1591

Qy    361 caaaggagcgtcgtcagcactggagcgccggccacgacgaggtctacggcggtgacgtcc 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1592 CAAAGGAGCGTCGTCAGCACTGGAGCGCCGGCCACGACGAGGTCTACGGCGGTGACGTCC 1651

Qy    421 acgactaccgccccgcagcgaacaagtccatacgggcacaatggccgacctgtgacggct 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1652 ACGACTACCGCCCCGCAGCGAACAAGTCCATACGGGCACAATGGCCGACCTGTGACGGCT 1711

Qy    481 ggattggtggcagctaatagtggtgccagcgcggcctcgtctcccacagccgcggcgaaa 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1712 GGATTGGTGGCAGCTAATAGTGGTGCCAGCGCGGCCTCGTCTCCCACAGCCGCGGCGAAA 1771

Qy    541 ccaacaggagaagaaaaggcctccgcggcatgtgaaacgagctccgtggcgata 594
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1772 CCAACAGGAGAAGAAAAGGCCTCCGCGGCATGTGAAACGAGCTCCGTGGCGATA 1825
```

RESULT 2
US-08-718-661-1
; Sequence 1, Application US/08718661
; Patent No. 5876972
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Nucleic acid molecules coding for mammalian
; TITLE OF INVENTION: tumor suppressor proteins and methods for their isolation
; NUMBER OF SEQUENCES: 15
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/718,661
; FILING DATE:
; CLASSIFICATION: 530
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2790 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 542..2545
US-08-718-661-1

Query Match 6.6%; Score 39.2; DB 2; Length 2790;
Best Local Similarity 54.1%; Pred. No. 0.093;
Matches 80; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

```
Qy    137 cggtagtggcgacggcgaaaccggggtgtgtgtcctcgggcactgacgcggcgagtagca 196
          | | ||  || || || |   | |         || | | ||||| | | || | |  ||
Db   2114 CAGCAGCAGCCACCGCCACTGCCGCCGCCACCACCGCTGCCACTGCCACAGCCACTGCCA 2173

Qy    197 gtcataccaatacgacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgc 256
          ||   ||    || ||  | ||| | | |||| || | |  |||    | | | |||| |
Db   2174 GTGCCACAGCCACTACCACAGCCACAGATGCAGCCACAGTTTCAGTTGCAGATCCAGCCC 2233

Qy    257 cagccaaggcagccgcgcctcctgccgc 284
          |||     ||  || |  || ||||||||
Db   2234 CAGATGCAGCTACCACAGCTGCTGCCGC 2261
```

RESULT 3
US-08-042-747A-7/c
; Sequence 7, Application US/08042747A
; Patent No. 5487969
; GENERAL INFORMATION:
; APPLICANT: Eberle, Richard
; APPLICANT: Black, Darla
; APPLICANT: Scinicariello, Franco
; APPLICANT: Hilliard, Julia K.
; TITLE OF INVENTION: Cloning and Amplification of Monkey B
; TITLE OF INVENTION: Virus Genes
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cox & Smith Incorporated
; STREET: 112 East Pecan Street, Suite 2000
; CITY: San Antonio
; STATE: Texas
; COUNTRY: USA
; ZIP: 78205
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/042,747A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Haymond, W. Bradley
; REGISTRATION NUMBER: 35186
; REFERENCE/DOCKET NUMBER: S-0072.179
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 210-554-5500
; TELEFAX: 210-226-8395
; TELEX: 767609
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2943 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 87..2744
US-08-042-747A-7

Query Match 6.4%; Score 38.2; DB 1; Length 2943;
Best Local Similarity 47.7%; Pred. No. 0.18;
Matches 112; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

```
Qy    209 cgacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaaggcag 268
          || | ||  |||   |  | ||  |  |   |  | | |     ||| | |    ||| |
Db    255 CGTCCTCGTCCGGGACCTCCTCGTCCTCGGGAAGCGCGGGCGTGGCGGCGGGGTCGGCGG 196

Qy    269 ccgcgcctcctgccgcggcgcgatcggctgagccgcacgtggggagcaagatcattgcta 328
```

```
           || | |  | ||||| ||| |  |||||   |||  || || || || |   |  || |
Db    195 TCGGGGCCGCCGCCGCCGCGGGGGCGGCTCGTCCGGCCGCGGCGATCACGGCGAGGGCGA 136

Qy    329 atctagtgaatcagctggggattaatgtcacccaaaggagcgtcgtcagcactggagcgc 388
              |     |    | |   | |          | | | |  || | | ||
Db    135 GGAGGACGGGGAGGGGAAGAAAGGAGGGGGGGGTGCCGCGAGGCCGCATCCCGGGGCGCG 76

Qy    389 cggccacgacgaggtctacggcggtgacgtccacgactaccgccccgcagcgaac 443
          ||| | | ||| || | ||  ||  | |||| ||  |  ||  || || |||| |
Db     75 CGGGCGCTACGCGGGCGACTACGCGGGCGTCGACTCCGGCCCGCCGGCCGCGAGC 21


RESULT   4
US-08-474-379C-60
; Sequence 60, Application US/08474379C
; Patent No. 5977305
;  GENERAL INFORMATION:
;    APPLICANT:  Wigler, Michael H.
;    APPLICANT:  Colicelli, John J.
;    TITLE OF INVENTION:  CLONING BY COMPLEMENTATION AND RELATED
;    TITLE OF INVENTION:  PROCESSES
;    NUMBER OF SEQUENCES:  88
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  233 South Wacker Drive/6300 Sears Tower
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/474,379C
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/511,715
;      FILING DATE:  20-APR-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/206,188
;      FILING DATE:  01-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/688,352
;      FILING DATE:  19-APR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  27866/32771
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 474-6300
;      TELEFAX:  (312) 474-0448
;  INFORMATION FOR SEQ ID NO:  60:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2115 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  762..2115
US-08-474-379C-60


  Query Match             6.3%;  Score 37.6;  DB 2;  Length 2115;
  Best Local Similarity   63.0%;  Pred. No. 0.23;
  Matches   58;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy    212 cttctgccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaaggcagccg 271
          || |  |||| || ||  |||  ||  || |||| |   ||| | |||||   || ||||
```

```
Db    577 CTGCCCCCGCCGCCGCCACCCTCGCCCCAGCCCCAGCCCCAGTGTCAGCCGCCGCCGCCG 636

Qy    272 cgcctcctgccgcggcgcgatcggctgagccg 303
          | || |||||||| || |   | || |  | |
Db    637 CCCCCCCTGCCGCCGCCCCCGCCGCCGCCCGG 668


RESULT   5
US-09-146-249A-60
; Sequence 60, Application US/09146249A
; Patent No. 6069240
;  GENERAL INFORMATION:
;    APPLICANT:  Wigler, Michael H.
;    APPLICANT:  Colicelli, John J.
;    TITLE OF INVENTION:  Cloning by Complementation and Related
;    TITLE OF INVENTION:  Processes
;    NUMBER OF SEQUENCES:  85
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/146,249A
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/511,715
;      FILING DATE:  20-APR-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312-474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  60:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2115 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  762..2115
US-09-146-249A-60


  Query Match             6.3%;  Score 37.6;  DB 3;  Length 2115;
  Best Local Similarity   63.0%;  Pred. No. 0.23;
  Matches   58;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy    212 cttctgccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaaggcagccg 271
          || |  |||| || ||  |||  ||  || |||| |   ||| | |||||   || ||||
Db    577 CTGCCCCCGCCGCCGCCACCCTCGCCCCAGCCCCAGCCCCAGTGTCAGCCGCCGCCGCCG 636

Qy    272 cgcctcctgccgcggcgcgatcggctgagccg 303
          | || |||||||| || |   | || |  | |
Db    637 CCCCCCCTGCCGCCGCCCCCGCCGCCGCCCGG 668


RESULT   6
US-08-206-188B-60
; Sequence 60, Application US/08206188B
; Patent No. 6100025
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  Wigler, Michael H.
;    APPLICANT:  Colicelli, John J.
;    TITLE OF INVENTION:  Cloning by Complementation and Related
;    TITLE OF INVENTION:  Processes
;    NUMBER OF SEQUENCES:  84
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/206,188B
;      FILING DATE:  01-MAR-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/511,715
;      FILING DATE:  20-APR-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36107
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312-474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  60:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2115 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  762..2115
US-08-206-188B-60


  Query Match             6.3%;  Score 37.6;  DB 3;  Length 2115;
  Best Local Similarity   63.0%;  Pred. No. 0.23;
  Matches   58;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy    212 cttctgccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaaggcagccg 271
          || |  |||| || ||  |||  ||  || |||| |   ||| | |||||   || ||||
Db    577 CTGCCCCCGCCGCCGCCACCCTCGCCCCAGCCCCAGCCCCAGTGTCAGCCGCCGCCGCCG 636

Qy    272 cgcctcctgccgcggcgcgatcggctgagccg 303
          | || |||||||| || |   | || |  | |
Db    637 CCCCCCCTGCCGCCGCCCCCGCCGCCGCCCGG 668


RESULT   7
US-09-121-321-15
; Sequence 15, Application US/09121321
; Patent No. 6090783
;  GENERAL INFORMATION:
;    APPLICANT:  Saiga, Akihiko
;    APPLICANT:  Orita, Satoshi
;    APPLICANT:  Igarashi, Hisanaga
;    APPLICANT:  Okumura, Kouichi
;    APPLICANT:  Sakaguchi, Gaku
;    TITLE OF INVENTION:  DNA MOLECULE RELATING TO SUPPRESSION OF
;    TITLE OF INVENTION:  GENE EXPRESSION AND NOVEL PROTEIN
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FISH & NEAVE
;      STREET:  1251 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10020
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/121,321
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/933,803
;      FILING DATE:  19-SEP-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Haley, James F.
;      REGISTRATION NUMBER:  27,794
;      REFERENCE/DOCKET NUMBER:  SHGN-12CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-596-9000
;      TELEFAX:  212-596-9090
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3777 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  139..2151
US-09-121-321-15


  Query Match             6.2%;  Score 36.8;  DB 3;  Length 3777;
  Best Local Similarity   69.4%;  Pred. No. 0.46;
  Matches   50;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy    231 ccctgcatcacccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcg 290
          ||| ||  || |||| |   ||||||||||  ||||||||  || | | ||| || |
Db   1355 CCCCGCCACAGCCCCAGCCCCAGCGCCAGCCACCGCAGCCGCAGCTGCAGTCGCAGCCCC 1414

Qy    291 atcggctgagcc 302
          | |  | |||||
Db   1415 AGCCCCAGAGCC 1426


RESULT   8
US-08-933-803A-15
; Sequence 15, Application US/08933803A
; Patent No. 6218522
;  GENERAL INFORMATION:
;    APPLICANT:  Saiga, Akihiko
;    APPLICANT:  Orita, Satoshi
;    APPLICANT:  Igarashi, Hisanaga
;    APPLICANT:  Okumura, Kouichi
;    APPLICANT:  Sakaguchi, Gaku
;    TITLE OF INVENTION:  DNA MOLECULE RELATING TO SUPPRESSION OF
;    TITLE OF INVENTION:  GENE EXPRESSION AND NOVEL PROTEIN
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FISH & NEAVE
;      STREET:  1251 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10020
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
```

```
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/933,803A
;       FILING DATE:  19-SEP-1997
;       CLASSIFICATION:  536
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Haley, James F.
;       REGISTRATION NUMBER:  27,794
;       REFERENCE/DOCKET NUMBER:  SHGN-12CIP
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  212-596-9000
;       TELEFAX:  212-596-9090
;   INFORMATION FOR SEQ ID NO:  15:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3777 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  139..2151
US-08-933-803A-15


  Query Match             6.2%;  Score 36.8;  DB 4;  Length 3777;
  Best Local Similarity   69.4%;  Pred. No. 0.46;
  Matches   50;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy    231 ccctgcatcacccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcg 290
          ||| ||  || |||| |   ||||||||||||  ||||||||  || | | ||| || |
Db   1355 CCCCGCCACAGCCCCAGCCCCAGCGCCAGCCACCGCAGCCGCAGCTGCAGTCGCAGCCCC 1414

Qy    291 atcggctgagcc 302
          | |  | |||||
Db   1415 AGCCCCAGAGCC 1426


RESULT  9
US-09-043-303-7/c
; Sequence 7, Application US/09043303
; Patent No. 6251589
; GENERAL INFORMATION:
;  APPLICANT: TSUJI, Shoji
;  APPLICANT: SANPEI, Kazujiro
;  TITLE OF INVENTION: Method for Diagnosing Spinocerebellar Alaxia Type 2 and
;  TITLE OF INVENTION: Primers Therefor
;  FILE REFERENCE: 0760-0241P
;  CURRENT APPLICATION NUMBER: US/09/043,303
;  CURRENT FILING DATE: 1998-05-18
;  EARLIER APPLICATION NUMBER: PCT/JP96/01999
;  EARLIER FILING DATE: 1996-07-18
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
;   LENGTH: 203
;   TYPE: DNA
;   ORGANISM: p-2093 plasmid
US-09-043-303-7


  Query Match             6.2%;  Score 36.6;  DB 4;  Length 203;
  Best Local Similarity   57.4%;  Pred. No. 0.21;
  Matches   66;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy    184 gcggcgagtagcagtcataccaatacgacttctgccgctgctgcatcccctgcatcaccc 243
          ||   ||||  | || || |   | |  || |||| ||||||||  |  ||||  |  |
Db    202 GCCCAGAGTTTCCGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 143

Qy    244 cctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggctg 298
          ||| | | ||  | ||   || || ||  || |||| || ||     | ||||
Db    142 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 88
```

```
RESULT  10
US-08-217-327-3
; Sequence 3, Application US/08217327
; Patent No. 5474925
;  GENERAL INFORMATION:
;    APPLICANT:  John, Maliyakal E
;    APPLICANT:  Barton, Kenneth A
;    TITLE OF INVENTION:  Immobilized Proteins in Cotton Fiber
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Quarles and Brady
;      STREET:  P.O. Box 2113
;      CITY:  Madison
;      STATE:  WI
;      COUNTRY:  USA
;      ZIP:  53701-2113
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/217,327
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/812,233
;      FILING DATE:  19-DEC-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seay, Nicholas J
;      REGISTRATION NUMBER:  27,386
;      REFERENCE/DOCKET NUMBER:  1122990831
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  608-251-5000
;      TELEFAX:  608-251-9166
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  913 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Gossypium hirsutum
;      STRAIN:  Coker 312
;      TISSUE TYPE:  Fiber cells
;    IMMEDIATE SOURCE:
;      CLONE:  H6
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  72..716
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  72..147
US-08-217-327-3


  Query Match             6.2%;  Score 36.6;  DB 1;  Length 913;
  Best Local Similarity   58.9%;  Pred. No. 0.33;
  Matches   63;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy    198 tcataccaatacgacttctgccgctgctgcatcccctgcatcaccccctgtttcagcgcc 257
          ||   ||| | |  || || | ||  || || | ||||| || || |||  | || | ||
Db    320 TCCACCCAGTTCTCCTCCTCCTGCAACTCCACCACCTGCTTCTCCTCCTCCTGCAACTCC 379

Qy    258 agccaaggcagccgcgcctcctgccgcggcgcgatcggctgagccgc 304
          | |    ||  |  | |||||||||| |  | | | | |||   || |
Db    380 ACCTCCAGCTTCTCCACCTCCTGCCACTCCTCCACCAGCTTCTCCAC 426
```

```
RESULT  11
US-07-885-970A-3
; Sequence 3, Application US/07885970A
; Patent No. 5495070
;  GENERAL INFORMATION:
;    APPLICANT:  John, Maliyakal E.
;    TITLE OF INVENTION:  GENETICALLY ENGINEERING COTTON
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nicholas J. Seay, Quarles & Brady
;      STREET:  P.O. Box 2113, First Wisconsin Plaza
;      CITY:  Madison
;      STATE:  Wisconsin
;      COUNTRY:  USA
;      ZIP:  53701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Microsoft Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/885,970A
;      FILING DATE:  19920518
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/617,239
;      FILING DATE:  21-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/253,243
;      FILING DATE:  04-OCT-1988
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seay, Nicholas J.
;      REGISTRATION NUMBER:  27,386
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (608) 283-2478
;      TELEFAX:  (608) 251-5139
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  913 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Gossypium hirsutum
;      STRAIN:  Coker 312
;      DEVELOPMENTAL STAGE:  15 day old fiber cells
;      TISSUE TYPE:  fiber cells
;    IMMEDIATE SOURCE:
;      LIBRARY:  CKFB15A1
;      CLONE:  H6
US-07-885-970A-3


  Query Match             6.2%;  Score 36.6;  DB 1;  Length 913;
  Best Local Similarity   58.9%;  Pred. No. 0.33;
  Matches   63;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy    198 tcataccaatacgacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgcc 257
          ||   ||| | |  || || | ||  || || | ||||| || || |||  | || | ||
Db    320 TCCACCCAGTTCTCCTCCTCCTGCAACTCCACCACCTGCTTCTCCTCCTCCTGCAACTCC 379

Qy    258 agccaaggcagccgcgcctcctgccgcggcgcgatcggctgagccgc 304
          | |    ||  |  | ||||||||| |  | | | | |||   || |
Db    380 ACCTCCAGCTTCTCCACCTCCTGCCACTCCTCCACCAGCTTCTCCAC 426


RESULT  12
US-08-298-687A-3
; Sequence 3, Application US/08298687A
; Patent No. 5521078
;  GENERAL INFORMATION:
;    APPLICANT:  John, Maliyakal E.
;    TITLE OF INVENTION:  GENETICALLY ENGINEERING COTTON
;    TITLE OF INVENTION:  PLANTS FOR ALTERED FIBER
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nicholas J. Seay, Quarles & Brady
;      STREET:  P.O. Box 2113, First Wisconsin Plaza
;      CITY:  Madison
;      STATE:  Wisconsin
;      COUNTRY:  USA
;      ZIP:  53701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Microsoft Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/298,687A
;      FILING DATE:
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/617,239
;      FILING DATE:  21-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/253,243
;      FILING DATE:  04-OCT-1988
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seay, Nicholas J.
;      REGISTRATION NUMBER:  27,386
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (608) 283-2478
;      TELEFAX:  (608) 251-5139
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  913 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Gossypium hirsutum
;      STRAIN:  Coker 312
;      DEVELOPMENTAL STAGE:  15 day old fiber cells
;      TISSUE TYPE:  fiber cells
;    IMMEDIATE SOURCE:
;      LIBRARY:  CKFB15A1
;      CLONE:  H6
US-08-298-687A-3


  Query Match             6.2%;  Score 36.6;  DB 1;  Length 913;
  Best Local Similarity   58.9%;  Pred. No. 0.33;
  Matches   63;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy    198 tcataccaatacgacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgcc 257
          ||   ||| | |  || || | ||  || || | ||||| || || |||  | || | ||
Db    320 TCCACCCAGTTCTCCTCCTCCTGCAACTCCACCACCTGCTTCTCCTCCTCCTGCAACTCC 379

Qy    258 agccaaggcagccgcgcctcctgccgcggcgcgatcggctgagccgc 304
          | |    ||  |  | ||||||||| |  | | | | |||   || |
Db    380 ACCTCCAGCTTCTCCACCTCCTGCCACTCCTCCACCAGCTTCTCCAC 426


RESULT  13
US-08-530-797-2
; Sequence 2, Application US/08530797
; Patent No. 5597718
;  GENERAL INFORMATION:
;    APPLICANT:  John, Maliyakal E.
```

```
;    APPLICANT:  Umbeck, Paul F.
;    APPLICANT:  Brill, Winston J.
;    TITLE OF INVENTION:  GENETICALY ENGINEERED COTTON PLANTS
;    TITLE OF INVENTION:  FOR ALTERED FIBER
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Quarles and Brady
;      STREET:  P.O BOX 2113
;      STREET:  FIRST WISCONSIN PLAZA
;      CITY:  MADISON
;      STATE:  WISCONSIN
;      COUNTRY:  U.S.A.
;      ZIP:  53701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette - 3.50 inch, 800Kb storage
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Macintosh
;      SOFTWARE:  Microsoft Word 4.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/530,797
;      FILING DATE:  20-SEP-1995
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/617,239
;      FILING DATE:  21-NOV-90
;      APPLICATION NUMBER:  US 07/253,243
;      FILING DATE:  04-OCT-88
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nicholas J. Seay
;      REGISTRATION NUMBER:  27,386
;      REFERENCE/DOCKET NUMBER:  1122990245
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  913 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  no
;    ANTI-SENSE:  no
;    ORIGINAL SOURCE:
;      ORGANISM:  Gossypium hirsutum
;      STRAIN:  Coker 312
;      DEVELOPMENTAL STAGE:  15 day old fiber cells
;      TISSUE TYPE:  fiber cells
;    IMMEDIATE SOURCE:
;      LIBRARY:  CKFB15A1
;      CLONE:  H6
US-08-530-797-2


  Query Match             6.2%;  Score 36.6;  DB 1;  Length 913;
  Best Local Similarity   58.9%;  Pred. No. 0.33;
  Matches   63;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy    198 tcataccaatacgacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgcc 257
          ||   ||| | |  || || | ||  || || | ||||| || || |||  | || | ||
Db    320 TCCACCCAGTTCTCCTCCTCCTGCAACTCCACCACCTGCTTCTCCTCCTCCTGCAACTCC 379

Qy    258 agccaaggcagccgcgcctcctgccgcggcgcgatcggctgagccgc 304
          | |    ||  |  | ||||||||| |  | | | | |||   || |
Db    380 ACCTCCAGCTTCTCCACCTCCTGCCACTCCTCCACCAGCTTCTCCAC 426


RESULT  14
US-08-298-829-3
; Sequence 3, Application US/08298829
; Patent No. 5620882
;  GENERAL INFORMATION:
;    APPLICANT:  John, Maliyakal E.
;    TITLE OF INVENTION:  GENETICALLY ENGINEERING COTTON
;    TITLE OF INVENTION:  PLANTS FOR ALTERED FIBER
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nicholas J. Seay, Quarles & Brady
;      STREET:  P.O. Box 2113, First Wisconsin Plaza
;      CITY:  Madison
;      STATE:  Wisconsin
;      COUNTRY:  USA
;      ZIP:  53701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Microsoft Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/298,829
;      FILING DATE:  19-OCT-1994
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/885,970
;      FILING DATE:  18-MAY-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/617,239
;      FILING DATE:  21-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/253,243
;      FILING DATE:  04-OCT-1988
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seay, Nicholas J.
;      REGISTRATION NUMBER:  27,386
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (608) 283-2478
;      TELEFAX:  (608) 251-5139
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  913 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Gossypium hirsutum
;      STRAIN:  Coker 312
;      DEVELOPMENTAL STAGE:  15 day old fiber cells
;      TISSUE TYPE:  fiber cells
;    IMMEDIATE SOURCE:
;      LIBRARY:  CKFB15A1
;      CLONE:  H6
US-08-298-829-3


  Query Match             6.2%;  Score 36.6;  DB 1;  Length 913;
  Best Local Similarity   58.9%;  Pred. No. 0.33;
  Matches   63;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy    198 tcataccaatacgacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgcc 257
          ||   ||| | |  || || | ||  || || | ||||| || || |||  | || | ||
Db    320 TCCACCCAGTTCTCCTCCTCCTGCAACTCCACCACCTGCTTCTCCTCCTCCTGCAACTCC 379

Qy    258 agccaaggcagccgcgcctcctgccgcggcgcgatcggctgagccgc 304
          | |    ||  |  | ||||||||| |  | | | | |||   || |
Db    380 ACCTCCAGCTTCTCCACCTCCTGCCACTCCTCCACCAGCTTCTCCAC 426


RESULT  15
US-08-787-335-2
; Sequence 2, Application US/08787335
; Patent No. 5981834
;  GENERAL INFORMATION:
;    APPLICANT:  John, Maliyakal E.
;    APPLICANT:  Umbeck, Paul F.
;    APPLICANT:  Brill, Winston J.
;    TITLE OF INVENTION:  GENETICALY ENGINEERED COTTON PLANTS
```

```
;    TITLE OF INVENTION:  FOR ALTERED FIBER
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Quarles and Brady
;      STREET:  P.O BOX 2113
;      STREET:  FIRST WISCONSIN PLAZA
;      CITY:  MADISON
;      STATE:  WISCONSIN
;      COUNTRY:  U.S.A.
;      ZIP:  53701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette - 3.50 inch, 800Kb storage
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Macintosh
;      SOFTWARE:  Microsoft Word 4.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/787,335
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/530,797
;      FILING DATE:
;      APPLICATION NUMBER:  US 07/253,243
;      FILING DATE:  04-OCT-88
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nicholas J. Seay
;      REGISTRATION NUMBER:  27,386
;      REFERENCE/DOCKET NUMBER:  1122990245
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  913 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  no
;    ANTI-SENSE:  no
;    ORIGINAL SOURCE:
;      ORGANISM:  Gossypium hirsutum
;      STRAIN:  Coker 312
;      DEVELOPMENTAL STAGE:  15 day old fiber cells
;      TISSUE TYPE:  fiber cells
;    IMMEDIATE SOURCE:
;      LIBRARY:  CKFB15A1
;      CLONE:  H6
US-08-787-335-2


  Query Match             6.2%;  Score 36.6;  DB 2;  Length 913;
  Best Local Similarity   58.9%;  Pred. No. 0.33;
  Matches   63;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy    198 tcataccaatacgacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgcc 257
          ||   ||| | |  || || | ||  || || | ||||| || || |||  | || | ||
Db    320 TCCACCCAGTTCTCCTCCTCCTGCAACTCCACCACCTGCTTCTCCTCCTCCTGCAACTCC 379

Qy    258 agccaaggcagccgcgcctcctgccgcggcgcgatcggctgagccgc 304
          | |    ||  |  | |||||||||  |  | | | | |||   || |  .
Db    380 ACCTCCAGCTTCTCCACCTCCTGCCACTCCTCCACCAGCTTCTCCAC 426


Search completed: September 21, 2001, 02:16:14
Job time: 28207 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: September 20, 2001, 17:52:08 ; Search time 6407.28 Seconds
(without alignments)
1369.914 Million cell updates/sec

Title: US-09-138-735-1_COPY_1232_1825
Perfect score: 594
Sequence: 1 caggtacagcgtaacggctt.....aaacgagctccgtggcgata 594

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 14155048 seqs, 7388405095 residues

Total number of hits satisfying chosen parameters: 28310096

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*
1: /cgn1_7/ptodata/1/pna/PCTUS_COMB.seq:*
2: /cgn1_7/ptodata/1/pna/US06_COMB.seq:*
3: /cgn1_7/ptodata/1/pna/US07_COMB.seq:*
4: /cgn1_7/ptodata/1/pna/US080_COMB.seq:*
5: /cgn1_7/ptodata/1/pna/US081_COMB.seq:*
6: /cgn1_7/ptodata/1/pna/US082_COMB.seq:*
7: /cgn1_7/ptodata/1/pna/US083_COMB.seq:*
8: /cgn1_7/ptodata/1/pna/US084_COMB.seq:*
9: /cgn1_7/ptodata/1/pna/US085_COMB.seq:*
10: /cgn1_7/ptodata/1/pna/US086_COMB.seq:*
11: /cgn1_7/ptodata/1/pna/US087_COMB.seq:*
12: /cgn1_7/ptodata/1/pna/US088_COMB.seq:*
13: /cgn1_7/ptodata/1/pna/US089_COMB.seq:*
14: /cgn1_7/ptodata/1/pna/US090_COMB.seq:*
15: /cgn1_7/ptodata/1/pna/US091_COMB.seq:*
16: /cgn1_7/ptodata/1/pna/US092_COMB.seq:*
17: /cgn1_7/ptodata/1/pna/US093_COMB.seq:*
18: /cgn1_7/ptodata/1/pna/US094_COMB.seq:*
19: /cgn1_7/ptodata/1/pna/US095A_COMB.seq:*
20: /cgn1_7/ptodata/1/pna/US095B_COMB.seq:*
21: /cgn1_7/ptodata/1/pna/US095C_COMB.seq:*
22: /cgn1_7/ptodata/1/pna/US095D_COMB.seq:*
23: /cgn1_7/ptodata/1/pna/US096A_COMB.seq:*
24: /cgn1_7/ptodata/1/pna/US096B_COMB.seq:*
25: /cgn1_7/ptodata/1/pna/US096C_COMB.seq:*
26: /cgn1_7/ptodata/1/pna/US096D_COMB.seq:*
27: /cgn1_7/ptodata/1/pna/US096E_COMB.seq:*
28: /cgn1_7/ptodata/1/pna/US097A_COMB.seq:*
29: /cgn1_7/ptodata/1/pna/US097B_COMB.seq:*
30: /cgn1_7/ptodata/1/pna/US097C_COMB.seq:*
31: /cgn1_7/ptodata/1/pna/US098_COMB.seq:*
32: /cgn1_7/ptodata/1/pna/US6000_COMB.seq:*
33: /cgn1_7/ptodata/1/pna/US6001_COMB.seq:*
34: /cgn1_7/ptodata/1/pna/US6002_COMB.seq:*
35: /cgn1_7/ptodata/1/pna/US6003_COMB.seq:*
36: /cgn1_7/ptodata/1/pna/US6004_COMB.seq:*
37: /cgn1_7/ptodata/1/pna/US6005_COMB.seq:*
38: /cgn1_7/ptodata/1/pna/US6006_COMB.seq:*
39: /cgn1_7/ptodata/1/pna/US6007_COMB.seq:*
40: /cgn1_7/ptodata/1/pna/US6008_COMB.seq:*
41: /cgn1_7/ptodata/1/pna/US6009_COMB.seq:*
42: /cgn1_7/ptodata/1/pna/US6010_COMB.seq:*
43: /cgn1_7/ptodata/1/pna/US6011_COMB.seq:*
44: /cgn1_7/ptodata/1/pna/US6012_COMB.seq:*
45: /cgn1_7/ptodata/1/pna/US6013_COMB.seq:*
46: /cgn1_7/ptodata/1/pna/US6014_COMB.seq:*
47: /cgn1_7/ptodata/1/pna/US6015_COMB.seq:*
48: /cgn1_7/ptodata/1/pna/US6016_COMB.seq:*
49: /cgn1_7/ptodata/1/pna/US6017_COMB.seq:*
50: /cgn1_7/ptodata/1/pna/US6018_COMB.seq:*
51: /cgn1_7/ptodata/1/pna/US6019_COMB.seq:*
52: /cgn1_7/ptodata/1/pna/US6020_COMB.seq:*
53: /cgn1_7/ptodata/1/pna/US6021_COMB.seq:*
54: /cgn1_7/ptodata/1/pna/US6022_COMB.seq:*
55: /cgn1_7/ptodata/1/pna/US6023_COMB.seq:*
56: /cgn1_7/ptodata/1/pna/US6024_COMB.seq:*
57: /cgn1_7/ptodata/1/pna/US6025_COMB.seq:*
58: /cgn1_7/ptodata/1/pna/US6026_COMB.seq:*
59: /cgn1_7/ptodata/1/pna/US6027_COMB.seq:*
60: /cgn1_7/ptodata/1/pna/US6028_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	594	100.0	3402	13	US-08-988-242-1	Sequence 1, Appli
	2	594	100.0	3402	15	US-09-138-736-1	Sequence 1, Appli
c	3	46	7.7	606	27	US-09-699-999-5885	Sequence 5885, Ap
c	4	45.8	7.7	470	27	US-09-699-999-4102	Sequence 4102, Ap
c	5	44.6	7.5	599	17	US-09-396-087-4576	Sequence 4576, App
	6	42.6	7.2	6257	24	US-09-620-392-44819	Sequence 44819, A
	7	42.6	7.2	29301	28	US-09-702-134-5721	Sequence 5721, Ap
	8	42.4	7.1	600	52	US-60-209-830-45296	Sequence 45296, A
c	9	42.2	7.1	491	51	US-60-197-872-57273	Sequence 57273, A
c	10	42.2	7.1	8953	24	US-09-620-392-68153	Sequence 68153, A
	11	42.2	7.1	10344	24	US-09-620-392-65686	Sequence 65686, A
c	12	42.2	7.1	10348	28	US-09-702-134-8344	Sequence 8344, Ap
	13	42	7.1	7227	48	US-60-168-139-2273	Sequence 2273, Ap
c	14	42	7.1	26276	19	US-09-514-000-337	Sequence 337, App
	15	42	7.1	397803	29	US-09-739-449-217	Sequence 217, App
	16	41.4	7.0	602	52	US-60-209-830-19436	Sequence 19436, A
c	17	41.2	6.9	12871	24	US-09-620-392-3460	Sequence 3460, Ap
c	18	41.2	6.9	33579	28	US-09-702-134-4946	Sequence 4946, Ap
	19	40.8	6.9	415	52	US-60-207-458-150303	Sequence 150303,
	20	40.8	6.9	547	52	US-60-207-458-149725	Sequence 149725,
	21	40.8	6.9	573	22	US-09-553-094-5427	Sequence 5427, Ap
	22	40.8	6.9	635	52	US-60-209-830-60545	Sequence 60545, A
	23	40.8	6.9	637	52	US-60-207-458-151348	Sequence 151348,
	24	40.8	6.9	668	52	US-60-207-458-151301	Sequence 151301,
c	25	40.6	6.8	550	51	US-60-197-872-44492	Sequence 44492, A
	26	40.4	6.8	470	25	US-09-654-617-111007	Sequence 111007,
	27	40.4	6.8	470	27	US-09-684-016-111007	Sequence 111007,
c	28	40.4	6.8	477	15	US-09-135-994-1	Sequence 1, Appli
	29	40.4	6.8	569	53	US-60-211-750-30151	Sequence 30151, A
	30	40.4	6.8	5537	1	PCT-US01-04098A-2481	Sequence 2481, Ap
	31	40.4	6.8	5537	18	US-09-496-914A-6707	Sequence 6707, Ap
	32	40.4	6.8	5537	22	US-09-560-875A-6707	Sequence 6707, Ap
c	33	40.4	6.8	6248	1	PCT-US01-04098A-513	Sequence 513, App
c	34	40.2	6.8	811	25	US-09-654-617-392521	Sequence 392521,
c	35	40.2	6.8	811	27	US-09-684-016-392521	Sequence 392521,
c	36	40	6.7	6867	24	US-09-620-392-66151	Sequence 66151, A
c	37	40	6.7	7443	24	US-09-620-392-55872	Sequence 55872, A
c	38	40	6.7	8299	24	US-09-620-392-34330	Sequence 34330, A
c	39	40	6.7	97314	24	US-09-620-392-30125	Sequence 30125, A
c	40	40	6.7	97370	28	US-09-702-134-10211	Sequence 10211, A
	41	39.8	6.7	431	17	US-09-394-745-23012	Sequence 23012, A
	42	39.8	6.7	431	22	US-09-565-306-47192	Sequence 47192, A
	43	39.8	6.7	1071	25	US-09-654-617-260110	Sequence 260110,
	44	39.8	6.7	1071	27	US-09-684-016-260110	Sequence 260110,
	45	39.8	6.7	16482	24	US-09-620-392-3109	Sequence 3109, Ap

ALIGNMENTS

```
RESULT   1
US-08-988-242-1
; Sequence 1, Application US/08988242
;  GENERAL INFORMATION:
;    APPLICANT:  PARANHOS-BACCALA, GLAUCIA
;    APPLICANT:  LESENECHAL, MYLENE
;    APPLICANT:  JOLIVET, MICHEL
;    APPLICANT:  MANDRAND, BERNARD
;    TITLE OF INVENTION:  NEW TRYPANOSOMA CRUZI ANTIGEN, GENE
;    TITLE OF INVENTION:  ENCODING THEREFOR, AND METHODS OF DETECTING AND TREATING
;    TITLE OF INVENTION:  CHAGAS DISEASE
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OLIFF & BERRIDGE, PLC
;      STREET:  P.O. BOX 19928
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22320
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/988,242
;      FILING DATE:  10-DEC-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berridge, William P.
;      REGISTRATION NUMBER:  30,024
;      REFERENCE/DOCKET NUMBER:  WPB 36400A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-836-6400
;      TELEFAX:  703-836-2787
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3402 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-988-242-1


  Query Match             100.0%;  Score 594;  DB 13;  Length 3402;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-126;
  Matches  594;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 caggtacagcgtaacggcttttgcttcaatcgtacagccgacggtagctgcgtcctggct 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1232 CAGGTACAGCGTAACGGCTTTTGCTTCAATCGTACAGCCGACGGTAGCTGCGTCCTGGCT 1291

Qy     61 gacatgtcgattcgattgacgatcttccatctccggtcctcccgcagggaagaacagcag 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1292 GACATGTCGATTCGATTGACGATCTTCCATCTCCGGTCCTCCCGCAGGGAAGAACAGCAG 1351

Qy    121 ccaggccaaaaaacatcggtagtggcgacggcgaaaccggggtgtgtgtcctcgggcact 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1352 CCAGGCCAAAAAACATCGGTAGTGGCGACGGCGAAACCGGGGTGTGTGTCCTCGGGCACT 1411

Qy    181 gacgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatcccctgcatca 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1412 GACGCGGCGAGTAGCAGTCATACCAATACGACTTCTGCCGCTGCTGCATCCCCTGCATCA 1471

Qy    241 ccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggctgag 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1472 CCCCCTGTTTCAGCGCCAGCCAAGGCAGCCGCGCCTCCTGCCGCGGCGCGATCGGCTGAG 1531

Qy    301 ccgcacgtggggagcaagatcattgctaatctagtgaatcagctggggattaatgtcacc 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1532 CCGCACGTGGGGAGCAAGATCATTGCTAATCTAGTGAATCAGCTGGGGATTAATGTCACC 1591

Qy    361 caaaggagcgtcgtcagcactggagcgccggccacgacgaggtctacggcggtgacgtcc 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1592 CAAAGGAGCGTCGTCAGCACTGGAGCGCCGGCCACGACGAGGTCTACGGCGGTGACGTCC 1651

Qy    421 acgactaccgccccgcagcgaacaagtccatacgggcacaatggccgacctgtgacggct 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1652 ACGACTACCGCCCCGCAGCGAACAAGTCCATACGGGCACAATGGCCGACCTGTGACGGCT 1711

Qy    481 ggattggtggcagctaatagtggtgccagcgcggcctcgtctcccacagccgcggcgaaa 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1712 GGATTGGTGGCAGCTAATAGTGGTGCCAGCGCGGCCTCGTCTCCCACAGCCGCGGCGAAA 1771

Qy    541 ccaacaggagaagaaaaggcctccgcggcatgtgaaacgagctccgtggcgata 594
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1772 CCAACAGGAGAAGAAAAGGCCTCCGCGGCATGTGAAACGAGCTCCGTGGCGATA 1825


RESULT   2
US-09-138-736-1
; Sequence 1, Application US/09138736
;  GENERAL INFORMATION:
;    APPLICANT:  PARANHOS-BACCALA, Glaucia
;    APPLICANT:  LESENECHAL, Mylene
;    APPLICANT:  JOLIVET, Michel
;    TITLE OF INVENTION:  NEW TRYPANOSOMA CRUZI ANTIGEN, AND GENE
;    TITLE OF INVENTION:  ENCODING THE LATTER; THEIR APPLICATION TO THE DETECTION OF
;    TITLE OF INVENTION:  CHAGAS DISEASE
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Oliff & Berridge
;      STREET:  700 South Washington Street, Suite 300
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22314
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/138,736
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/480,917
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berridge, William P.
;      REGISTRATION NUMBER:  30,024
;      REFERENCE/DOCKET NUMBER:  WPB 36400
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-836-6400
;      TELEFAX:  703-836-2787
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3402 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-138-736-1


  Query Match             100.0%;  Score 594;  DB 15;  Length 3402;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-126;
  Matches  594;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 caggtacagcgtaacggcttttgcttcaatcgtacagccgacggtagctgcgtcctggct 60
```

```
Db   1232 CAGGTACAGCGTAACGGCTTTTGCTTCAATCGTACAGCCGACGGTAGCTGCGTCCTGGCT 1291

Qy     61 gacatgtcgattcgattgacgatcttccatctccggtcctcccgcagggaagaacagcag 120
Db   1292 GACATGTCGATTCGATTGACGATCTTCCATCTCCGGTCCTCCCGCAGGGAAGAACAGCAG 1351

Qy    121 ccaggccaaaaaacatcggtagtggcgacggcgaaaccggggtgtgtgtcctcgggcact 180
Db   1352 CCAGGCCAAAAAACATCGGTAGTGGCGACGGCGAAACCGGGGTGTGTGTCCTCGGGCACT 1411

Qy    181 gacgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatcccctgcatca 240
Db   1412 GACGCGGCGAGTAGCAGTCATACCAATACGACTTCTGCCGCTGCTGCATCCCCTGCATCA 1471

Qy    241 ccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggctgag 300
Db   1472 CCCCCTGTTTCAGCGCCAGCCAAGGCAGCCGCGCCTCCTGCCGCGGCGCGATCGGCTGAG 1531

Qy    301 ccgcacgtggggagcaagatcattgctaatctagtgaatcagctggggattaatgtcacc 360
Db   1532 CCGCACGTGGGGAGCAAGATCATTGCTAATCTAGTGAATCAGCTGGGGATTAATGTCACC 1591

Qy    361 caaaggagcgtcgtcagcactggagcgccggccacgacgaggtctacggcggtgacgtcc 420
Db   1592 CAAAGGAGCGTCGTCAGCACTGGAGCGCCGGCCACGACGAGGTCTACGGCGGTGACGTCC 1651

Qy    421 acgactaccgccccgcagcgaacaagtccatacgggcacaatggccgacctgtgacggct 480
Db   1652 ACGACTACCGCCCCGCAGCGAACAAGTCCATACGGGCACAATGGCCGACCTGTGACGGCT 1711

Qy    481 ggattggtggcagctaatagtggtgccagcgcggcctcgtctcccacagccgcggcgaaa 540
Db   1712 GGATTGGTGGCAGCTAATAGTGGTGCCAGCGCGGCCTCGTCTCCCACAGCCGCGGCGAAA 1771

Qy    541 ccaacaggagaagaaaaggcctccgcggcatgtgaaacgagctccgtggcgata 594
Db   1772 CCAACAGGAGAAGAAAAGGCCTCCGCGGCATGTGAAACGAGCTCCGTGGCGATA 1825
```

RESULT 3
US-09-699-999-5885/c
; Sequence 5885, Application US/09699999
; GENERAL INFORMATION:
; APPLICANT: Shyjan, Andrew W.
; APPLICANT: MacBeth, Kyle J.
; APPLICANT: Vasicek, Thomas
; APPLICANT: Wang, Youzhan
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600.2032-001
; CURRENT APPLICATION NUMBER: US/09/699,999
; CURRENT FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: 60/162,361
; PRIOR FILING DATE: 1999-10-29
; NUMBER OF SEQ ID NOS: 7488
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5885
; LENGTH: 606
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(606)
; OTHER INFORMATION: n = A,T,C or G
US-09-699-999-5885

Query Match 7.7%; Score 46; DB 27; Length 606;
Best Local Similarity 46.5%; Pred. No. 1.2;
Matches 148; Conservative 0; Mismatches 170; Indels 0; Gaps 0;

```
Qy    106 agggaagaacagcagccaggccaaaaaacatcggtagtggcgacggcgaaaccggggtgt 165
Db    324 AGGGCTGAGGACCCCGGGGGCATGAACACGCCGGAGGGGGCGGCGGTGGGGGGCGGGCGG 265

Qy    166 gtgtcctcgggcactgacgcggcgagtagcagtcataccaatacgacttctgccgctgct 225
Db    264 CTTCCTTCAGGAGGGCGCGCCCGCCGAGCGCCGCTCTCCGCGGCGGCTGCTGCTGTGGCT 205

Qy    226 gcatcccctgcatcacccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcg 285
Db    204 GTCGCTGCTGCTGCCGCTCTAACTGCAGTGGCGGCTGCTGTGGCTGTGGCTGCTGCTGCT 145

Qy    286 gcgcgatcggctgagccgcacgtggggagcaagatcattgctaatctagtgaatcagctg 345
Db    144 CCCCGGGCCGCAGGCGCGTCCGCGTCGTGGCCGCTGCTGGGATCGCTGGTGCCGCCGCCG 85

Qy    346 gggattaatgtcacccaaaggagcgtcgtcagcactggagcgccggccacgacgaggtct 405
Db     84 CCGCCGCCAGGCGCCCGCTCGCATGTGGCTCGCGCCGCGTCTCAGGCCGCCCGGCGGATC 25

Qy    406 acggcggtgacgtccacg 423
Db     24 CCGGCGGACGCGTGGGCG 7
```

RESULT 4
US-09-699-999-4102/c
; Sequence 4102, Application US/09699999
; GENERAL INFORMATION:
; APPLICANT: Shyjan, Andrew W.
; APPLICANT: MacBeth, Kyle J.
; APPLICANT: Vasicek, Thomas
; APPLICANT: Wang, Youzhan
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600.2032-001
; CURRENT APPLICATION NUMBER: US/09/699,999
; CURRENT FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: 60/162,361
; PRIOR FILING DATE: 1999-10-29
; NUMBER OF SEQ ID NOS: 7488
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4102
; LENGTH: 470
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(470)
; OTHER INFORMATION: n = A,T,C or G
US-09-699-999-4102

Query Match 7.7%; Score 45.8; DB 27; Length 470;
Best Local Similarity 46.6%; Pred. No. 1.4;
Matches 146; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

```
Qy    106 agggaagaacagcagccaggccaaaaaacatcggtagtggcgacggcgaaaccggggtgt 165
Db    319 AGGGCTGAGGACCCCGGGGGCATGAACACGCCGGAGGGGGCGGCGGTGGGGGGCGGGCGG 260

Qy    166 gtgtcctcgggcactgacgcggcgagtagcagtcataccaatacgacttctgccgctgct 225
Db    259 CTTCCTTCAGGAGGGCGCGCCCGCCGAGCGCCGCTCTCCGCGGCGGCTGCTGCTGTGGCT 200

Qy    226 gcatcccctgcatcacccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcg 285
Db    199 GTCGCTGCTGCTGCCGCTCTAACTGCAGTGGCGGCTGCTGTGGCTGTGGCTGCTGCTGCT 140

Qy    286 gcgcgatcggctgagccgcacgtggggagcaagatcattgctaatctagtgaatcagctg 345
Db    139 CCCCGGGCCGCAGGCGCGTCCGCGTCGTGGCCGCTGCTGGGATCGCTGGTGCCGCCGCCG 80

Qy    346 gggattaatgtcacccaaaggagcgtcgtcagcactggagcgccggccacgacgaggtct 405
```

```
Db     79 CCGCCGCCAGGCGCCCGCTCGCATGTGGCTCGCGCCGCGTCTCAGGCCGCCCGGCGGATC 20

Qy    406 acggcggtgacgt 418
          |||||| |||
Db     19 CCGGCGGACGCGT 7


RESULT   5
US-09-396-087-4576/c
; Sequence 4576, Application US/09396087
; GENERAL INFORMATION:
;  APPLICANT: Gearing, David P.
;  APPLICANT: Holtzman, Douglas A.
;  TITLE OF INVENTION: NUCLEIC ACID MOLECULES DERIVED FROM A
;  TITLE OF INVENTION: PULMONARY ARTERY SMOOTH MUSCLE LIBRARY
;  FILE REFERENCE: MLN98-39pM
;  CURRENT APPLICATION NUMBER: US/09/396,087
;  CURRENT FILING DATE: 1999-09-14
;  EARLIER APPLICATION NUMBER: 60/100,260
;  EARLIER FILING DATE: 1998-09-14
;  EARLIER APPLICATION NUMBER: 60/107,226
;  EARLIER FILING DATE: 1998-11-05
;  EARLIER APPLICATION NUMBER: 60/131,810
;  EARLIER FILING DATE: 1999-04-30
;  NUMBER OF SEQ ID NOS: 5220
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4576
;   LENGTH: 599
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(599)
;   OTHER INFORMATION: n = A,T,C or G
US-09-396-087-4576


  Query Match             7.5%;  Score 44.6;  DB 17;  Length 599;
  Best Local Similarity   46.6%;  Pred. No. 2.6;
  Matches  143;  Conservative    0;  Mismatches  164;  Indels    0;  Gaps    0;

Qy    106 agggaagaacagcagccaggccaaaaaacatcggtagtggcgacggcgaaaccggggtgt 165
          ||||  ||  | |     |||   || ||  |||  | |||| ||| |      ||| |
Db    316 AGGGCTGAGGACCCCGGGGGCATGAACACGCCGGAGGGGGCGGCGGTGGGGGGCGGGCGG 257

Qy    166 gtgtcctcgggcactgacgcggcgagtagcagtcataccaatacgacttctgccgctgct 225
          |  | || ||      |||     |       |   ||    || || |||| |  |||
Db    256 CTTCCTTCAGGAGGGCGCGCCCGCCGAGCGCCGCTCTCCGCGGCGGCTGCTGCTGTGGCT 197

Qy    226 gcatcccctgcatcacccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcg 285
          |   |  |||| |  | |    | ||| | | ||    |  || | | || |||| ||
Db    196 GTCGCTGCTGCTGCCGCTCTAACTGCAGTGGCGGCTGCTGTGGCTGTGGCTGCTGCTGCT 137

Qy    286 gcgcgatcggctgagccgcacgtggggagcaagatcattgctaatctagtgaatcagctg 345
          | ||  | || |   ||  || | | |   | |  | |     || |||   | || |
Db    136 CCCCGGGCCGCAGGCGCGTCCGCGTCGTGGCCGCTGCTGGGATCGCTGGTGCCGCCGCCG 77

Qy    346 gggattaatgtcacccaaaggagcgtcgtcagcactggagcgccggccacgacgaggtct 405
          |      | | |||    |   || |   || | |   | | |||| |   | ||
Db     76 CCGCCGCCAGGCGCCCGCTCGCATGTGGCTCGCGCCGCGTCTCAGGCCGCCCGGCGGATC 17

Qy    406 acggcgg 412
          |||||||
Db     16 CCGGCGG 10


RESULT   6
US-09-620-392-44819
; Sequence 44819, Application US/09620392
; GENERAL INFORMATION:
;  APPLICANT:       Boukharov, Andrey A.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Liu, Jingdong
;  APPLICANT:  McIninch, James
;  TITLE OF INVENTION:       Plant Genome Sequence And Uses Thereof
;  FILE REFERENCE:       38-21(51237)E
;  CURRENT APPLICATION NUMBER: US/09/620,392
;  CURRENT FILING DATE:  2000-07-19
;  NUMBER OF SEQ ID NOS:       69652
; SEQ ID NO 44819
;   LENGTH: 6257
;   TYPE: DNA
;   ORGANISM: Oryza sativa
US-09-620-392-44819


  Query Match             7.2%;  Score 42.6;  DB 24;  Length 6257;
  Best Local Similarity   48.2%;  Pred. No. 8.8;
  Matches  120;  Conservative    0;  Mismatches  129;  Indels    0;  Gaps    0;

Qy    161 ggtgtgtgtcctcgggcactgacgcggcgagtagcagtcataccaatacgacttctgccg 220
          | |  | | || |||   | |  ||||      ||        ||   |||| | || |
Db   2350 gttcggcggccccggcggccgcggcggacccgcgctcggccctcacggcgacgacggcgg 2409

Qy    221 ctgctgcatcccctgcatcacccccctgtttcagcgccagccaaggcagccgcgcctcctg 280
          |  | ||  | || ||  |   | | |  || | | |  |  |||| || ||| |  | |
Db   2410 caccggcggcaccggcggctttctcggcgtcggaggcgacggaggccgcggcggcgacgg 2469

Qy    281 ccgcggcgcgatcggctgagccgcacgtggggagcaagatcattgctaatctagtgaatc 340
          | ||   |    | || ||| |    | || ||   || |||  ||  | || |  ||
Db   2470 cggcccaggaggcagcggaggccgcggcggcgacggaggtcaaggcggacctggccaacg 2529

Qy    341 agctggggattaatgtcacccaaaggagcgtcgtcagcactggagcgccggccacgacga 400
          |  |||||     | | |     ||   |  | |    | ||  || |||| |||  |
Db   2530 gggcggggacggcgggcccggcggggcaggcgggcccggcgggcacggcggcgacggtgg 2589

Qy    401 ggtctacgg 409
          || |  |||
Db   2590 ggacggcgg 2598


RESULT   7
US-09-702-134-5721
; Sequence 5721, Application US/09702134
; GENERAL INFORMATION:
;  APPLICANT:       Boukharov, Andrey A.
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Liu, Jingdong
;  APPLICANT:  McIninch, James
;  APPLICANT:  Wu, Wei
;  TITLE OF INVENTION:       Plant Genome Sequence and Uses Thereof
;  FILE REFERENCE:       38-21(51237)F
;  CURRENT APPLICATION NUMBER: US/09/702,134
;  CURRENT FILING DATE:  2000-10-31
;  NUMBER OF SEQ ID NOS:       52202
; SEQ ID NO 5721
;   LENGTH: 29301
;   TYPE: DNA
;   ORGANISM: Oryza sativa
US-09-702-134-5721


  Query Match             7.2%;  Score 42.6;  DB 28;  Length 29301;
  Best Local Similarity   48.2%;  Pred. No. 9.8;
  Matches  120;  Conservative    0;  Mismatches  129;  Indels    0;  Gaps    0;

Qy    161 ggtgtgtgtcctcgggcactgacgcggcgagtagcagtcataccaatacgacttctgccg 220
          | |  | | || |||   | |  ||||      ||        ||   |||| | || |
Db   2414 gttcggcggccccggcggccgcggcggacccgcgctcggccctcacggcgacgacggcgg 2473

Qy    221 ctgctgcatcccctgcatcacccccctgtttcagcgccagccaaggcagccgcgcctcctg 280
          |  | ||  | || ||  |   | | |  || | | |  |  |||| || ||| |  | |
Db   2474 caccggcggcaccggcggctttctcggcgtcggaggcgacggaggccgcggcggcgacgg 2533
```

```
Qy    281 ccgcggcgcgatcggctgagccgcacgtggggagcaagatcattgctaatctagtgaatc 340
          | ||   |    | || ||| |    | || ||   || |||  ||  | || |  ||
Db   2534 cggcccaggaggcagcggaggccgcggcggcgacggaggtcaaggcggacctggccaacg 2593

Qy    341 agctggggattaatgtcacccaaaggagcgtcgtcagcactggagcgccggccacgacga 400
          |  |||||     | | |     ||   |  | |    | ||  || |||| ||| |
Db   2594 gggcggggacggcgggcccggcggggcaggcgggcccggcgggcacggcggcgacggtgg 2653

Qy    401 ggtctacgg 409
          || |  |||
Db   2654 ggacggcgg 2662


RESULT   8
US-60-209-830-45296
; Sequence 45296, Application US/60209830
; GENERAL INFORMATION:
;  APPLICANT:       Andersen, Scott E.
;  APPLICANT:  Castiglioni, Paolo
;  APPLICANT:  Hardeman, Kristine J.
;  APPLICANT:  Kerk, Nancy M.
;  APPLICANT:  La Rosa, Thomas J.
;  TITLE OF INVENTION:       NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
;  TITLE OF INVENTION:  PLANTS
;  FILE REFERENCE:       38-21(51934)A
;  CURRENT APPLICATION NUMBER: US/60/209,830
;  CURRENT FILING DATE:  2000-06-06
;  NUMBER OF SEQ ID NOS:       62628
; SEQ ID NO 45296
;   LENGTH: 600
;   TYPE: DNA
;   ORGANISM: Zea mays
;   OTHER INFORMATION: Clone ID: LIB3596-038-Q6-K6-G7
US-60-209-830-45296


  Query Match             7.1%;  Score 42.4;  DB 52;  Length 600;
  Best Local Similarity   60.3%;  Pred. No. 8.4;
  Matches   70;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy    178 actgacgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatcccctgca 237
          || | | | |||||| || |||  |||  | | ||  | ||| || | ||  | ||| |
Db    399 acagccccagcgagtcgccgtcgcacccctcctacggccgccactcccgctcctcctccg 458

Qy    238 tcacccccgtgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgatc 293
          ||| || ||  | | ||| |  ||| ||  || ||| |  | | |||   | | ||
Db    459 tcagccgcttctccggcggcctccacgggggcggcggcggcgggcgcaaggtgctc 514


RESULT   9
US-60-197-872-57273/c
; Sequence 57273, Application US/60197872
; GENERAL INFORMATION:
;  APPLICANT:       Bougri, Olegs
;  APPLICANT:  Byrum, Joseph R.
;  APPLICANT:  De La Pena, Robert C.
;  APPLICANT:  La Rosa, Thomas J.
;  APPLICANT:  Shukla, Hridayabhiranjan
;  TITLE OF INVENTION:       Nucleic acid Molecules and Other molecules associated with
;  TITLE OF INVENTION:  plants
;  FILE REFERENCE:       38-21(51892)A
;  CURRENT APPLICATION NUMBER: US/60/197,872
;  CURRENT FILING DATE:  2000-04-19
;  NUMBER OF SEQ ID NOS:       76255
; SEQ ID NO 57273
;   LENGTH: 491
;   TYPE: DNA
;   ORGANISM: Oryza sativa nipponbare
;   OTHER INFORMATION: Clone ID: jC-osflLIB3479034a01b1
US-60-197-872-57273


  Query Match             7.1%;  Score 42.2;  DB 51;  Length 491;
  Best Local Similarity   53.3%;  Pred. No. 9.2;
  Matches   89;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy    142 gtggcgacggcgaaaccggggtgtgtgtcctcgggcactgacgcggcgagtagcagtcat 201
          | |||  || |     |||  ||||  ||| | || ||||  |   || ||  |||||
Db    371 GCGGCAGCGTCAGCGGCGGCTTGTGGTTCCGCAGGTACTGCAGGAACGTGTTCCAGTCCC 312

Qy    202 accaatacgacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgccagcc 261
          ||  | |  |  ||     |     || ||||  |  | ||  |   | | ||||||||
Db    311 GCCGCTTCTGCGACTCGTACCTGCTCAGCCCCGCCGCCGCCGGCGCATGCTGCGCCAGCG 252

Qy    262 aaggcagccgcgcctcctgccgcggcgcgatcggctgagccgcacgt 308
            || || |||| || |||  | ||     || ||| | |  | ||
Db    251 GCTGCGGCTGCGCTTCTTGCACCTGCTGCTGCGCCTGCGACTGATGT 205


RESULT  10
US-09-620-392-68153/c
; Sequence 68153, Application US/09620392
; GENERAL INFORMATION:
;  APPLICANT:       Boukharov, Andrey A.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Liu, Jingdong
;  APPLICANT:  McIninch, James
;  TITLE OF INVENTION:       Plant Genome Sequence And Uses Thereof
;  FILE REFERENCE:       38-21(51237)E
;  CURRENT APPLICATION NUMBER: US/09/620,392
;  CURRENT FILING DATE:  2000-07-19
;  NUMBER OF SEQ ID NOS:       69652
; SEQ ID NO 68153
;   LENGTH: 8953
;   TYPE: DNA
;   ORGANISM: Oryza sativa
US-09-620-392-68153


  Query Match             7.1%;  Score 42.2;  DB 24;  Length 8953;
  Best Local Similarity   51.9%;  Pred. No. 11;
  Matches   95;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy    133 acatcggtagtggcgacggcgaaaccggggtgtgtgtcctcgggcactgacgcggcgagt 192
          ||| | |    ||| |  |   ||||  | ||  | || |||    |    |||||  ||
Db   1548 ACACCCGCGCCTGCGCCTCCAGCACCGTCGCGTTCGGCCGCGGCATCCCCGGCGGCCGGT 1489

Qy    193 agcagtcataccaatacgacttctgccgctgctgcatcccctgcatcacccccctgtttca 252
          |  ||||  ||     ||  ||| ||| |  |  |  | || ||| | || ||     |
Db   1488 ACGAGTCCGACACCGCCGGGTTCCGCCTCACCCCCTCCTCCAGCACCGCCACCCCCGCCT 1429

Qy    253 gcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggctgagccgcacgtgggg 312
          | ||| ||| | |  |  |||| ||  || || ||| | |   |    || | ||  |
Db   1428 CCTCCATCCACGACGCCTCCGCCCCCGCCCCCGCCGCCACCACCGCCACCCCGCGCCGCC 1369

Qy    313 agc 315
          | |
Db   1368 ATC 1366


RESULT  11
US-09-620-392-65686
; Sequence 65686, Application US/09620392
; GENERAL INFORMATION:
;  APPLICANT:       Boukharov, Andrey A.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Liu, Jingdong
;  APPLICANT:  McIninch, James
;  TITLE OF INVENTION:       Plant Genome Sequence And Uses Thereof
;  FILE REFERENCE:       38-21(51237)E
;  CURRENT APPLICATION NUMBER: US/09/620,392
;  CURRENT FILING DATE:  2000-07-19
;  NUMBER OF SEQ ID NOS:       69652
; SEQ ID NO 65686
;   LENGTH: 10344
;   TYPE: DNA
```

```
;   ORGANISM: Oryza sativa
US-09-620-392-65686


  Query Match             7.1%;  Score 42.2;  DB 24;  Length 10344;
  Best Local Similarity   51.9%;  Pred. No. 11;
  Matches   95;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy     133 acatcggtagtggcgacggcgaaaccggggtgtgtgtcctcgggcactgacgcggcgagt 192
           ||| | |     ||| |  |   |||  | ||  | || |||   |    |||||  ||
Db    8556 acacccgcgcctgcgcctccagcaccgtcgcgttcggccgcggcatccccggcggccggt 8615

Qy     193 agcagtcataccaatacgacttctgccgctgctgcatcccctgcatcacccccctgtttca 252
           |  ||||  ||     ||  ||| ||| |  |  |  | || ||| | || ||     |
Db    8616 acgagtccgacaccgccgggttccgcctcacccccctcctccagcaccgccacccccgcct 8675

Qy     253 gcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggctgagccgcacgtgggg 312
           | ||| ||| | |  |  |||| ||  || || ||| | |  |    || | ||  |
Db    8676 cctccatccacgacgcctccgcccccgccccgccgccaccaccgccaccccgcgccgcc 8735

Qy     313 agc 315
           | |
Db    8736 atc 8738


RESULT  12
US-09-702-134-8344/c
; Sequence 8344, Application US/09702134
; GENERAL INFORMATION:
;  APPLICANT:       Boukharov, Andrey A.
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Liu, Jingdong
;  APPLICANT:  McIninch, James
;  APPLICANT:  Wu, Wei
;  TITLE OF INVENTION:       Plant Genome Sequence and Uses Thereof
;  FILE REFERENCE:       38-21(51237)F
;  CURRENT APPLICATION NUMBER: US/09/702,134
;  CURRENT FILING DATE:  2000-10-31
;  NUMBER OF SEQ ID NOS:       52202
; SEQ ID NO 8344
;   LENGTH: 10348
;   TYPE: DNA
;   ORGANISM: Oryza sativa
US-09-702-134-8344


  Query Match             7.1%;  Score 42.2;  DB 28;  Length 10348;
  Best Local Similarity   51.9%;  Pred. No. 11;
  Matches   95;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy     133 acatcggtagtggcgacggcgaaaccggggtgtgtgtcctcgggcactgacgcggcgagt 192
           ||| | |     ||| |  |   ||||  | ||  | || |||   |    |||||  ||
Db    1793 ACACCCGCGCCTGCGCCTCCAGCACCGTCGCGTTCGGCCGCGGCATCCCCGGCGGCCGGT 1734

Qy     193 agcagtcataccaatacgacttctgccgctgctgcatcccctgcatcacccccctgtttca 252
           |  ||||  ||     ||  ||| ||| |  |  |  | || ||| | || ||     |
Db    1733 ACGAGTCCGACACCGCCGGGTTCCGCCTCACCCCCTCCTCCAGCACCGCCACCCCCGCCT 1674

Qy     253 gcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggctgagccgcacgtgggg 312
           | ||| ||| | |  |  |||| ||  || || ||| | |  |    || | ||  |
Db    1673 CCTCCATCCACGACGCCTCCGCCCCCGCCCCCGCCGCCACCACCGCCACCCCGCGCCGCC 1614

Qy     313 agc 315
           | |
Db    1613 ATC 1611


RESULT  13
US-60-168-139-2273
; Sequence 2273, Application US/60168139
; GENERAL INFORMATION:
;  APPLICANT: Hinkle, Gregory J.
;  APPLICANT: Slater, Steven C.
;  TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequence And Uses Thereof
;  FILE REFERENCE: 38-21(15490)A
;  CURRENT APPLICATION NUMBER: US/60/168,139
;  CURRENT FILING DATE: 1999-12-01
;  NUMBER OF SEQ ID NOS: 3432
; SEQ ID NO 2273
;   LENGTH: 7227
;   TYPE: DNA
;   ORGANISM: Agrobacterium tumefaciens
;   FEATURE:
;   OTHER INFORMATION: unsure at all n locations
US-60-168-139-2273


  Query Match             7.1%;  Score 42;  DB 48;  Length 7227;
  Best Local Similarity   60.5%;  Pred. No. 12;
  Matches   69;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy     183 cgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatcccctgcatcacc 242
           | | || | || |  ||  |     ||  || | || || || ||| ||||| || ||||
Db    1696 ccccgccaatatcgttcccaaactgacacctcccgcggcggcggcagcccctccagcacc 1755

Qy     243 ccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggc 296
               ||| | || || ||   || |   || ||||| |||||||| | | | ||
Db    1756 ggaagttgctgcaccggcagcggtacttgctcctccagccgcggctccacccgc 1809


RESULT  14
US-09-514-000-337/c
; Sequence 337, Application US/09514000
; GENERAL INFORMATION:
;  APPLICANT: Hinkle, Gregory J.
;  APPLICANT: Slater, Steven C.
;  TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
;  FILE REFERENCE: 38-10(15490)B
;  CURRENT APPLICATION NUMBER: US/09/514,000
;  CURRENT FILING DATE: 2000-02-23
;  NUMBER OF SEQ ID NOS: 15034
; SEQ ID NO 337
;   LENGTH: 26276
;   TYPE: DNA
;   ORGANISM: Agrobacterium tumefaciens
US-09-514-000-337


  Query Match             7.1%;  Score 42;  DB 19;  Length 26276;
  Best Local Similarity   60.5%;  Pred. No. 13;
  Matches   69;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy     183 cgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatcccctgcatcacc 242
           | | || | || |  ||  |     ||  || | || || || ||| ||||| || ||||
Db   14789 CCCCGCCAATATCGTTCCCAAACTGACACCTCCCGCGGCGGCGGCAGCCCCTCCAGCACC 14730

Qy     243 ccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggc 296
               ||| | || || ||   || |   || ||||| |||||||| | | | ||
Db   14729 GGAAGTTGCTGCACCGGCAGCGGTACTTGCTCCTCCAGCCGCGGCTCCACCCGC 14676


RESULT  15
US-09-739-449-217
; Sequence 217, Application US/09739449
; GENERAL INFORMATION:
;  APPLICANT: Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
;  FILE REFERENCE: 38-10(15490)C
;  CURRENT APPLICATION NUMBER: US/09/739,449
;  CURRENT FILING DATE:  2000-12-19
;  PRIOR APPLICATION NUMBER: US 09/514,000
;  PRIOR FILING DATE: 2000-02-23
;  NUMBER OF SEQ ID NOS: 13351
; SEQ ID NO 217
```

; LENGTH: 397803
; TYPE: DNA
; ORGANISM: Agrobacterium tumefaciens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(397803)
; OTHER INFORMATION: unsure at all n locations
US-09-739-449-217

Query Match 7.1%; Score 42; DB 29; Length 397803;
Best Local Similarity 60.5%; Pred. No. 16;
Matches 69; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

```
Qy    183 cgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatcccctgcatcacc 242
          | | || | || |  ||  |     ||  || | || || || ||| ||||| || ||||
Db  66682 ccccgccaatatcgttcccaaactgacacctcccgcggcggcggcagcccctccagcacc 66741

Qy    243 ccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggc 296
              ||| | || || ||   || |   || ||||| |||||||| | | | ||
Db  66742 ggaagttgctgcaccggcagcggtacttgctcctccagccgcggctccacccgc 66795
```

Search completed: September 21, 2001, 02:15:37
Job time: 30209 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: September 20, 2001, 22:28:34 ; Search time 1035.34 Seconds
(without alignments)
1145.007 Million cell updates/sec

Title: US-09-138-735-1_COPY_1232_1825
Perfect score: 594
Sequence: 1 caggtacagcgtaacggctt.....aaacgagctccgtggcgata 594

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1741792 seqs, 997871973 residues

Total number of hits satisfying chosen parameters: 3483584

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_New:*
1: /cgn1_7/ptodata/2/pna/PCT_NEW_COMB.seq:*
2: /cgn1_7/ptodata/2/pna/US06_NEW_COMB.seq:*
3: /cgn1_7/ptodata/2/pna/US07_NEW_COMB.seq:*
4: /cgn1_7/ptodata/2/pna/US08_NEW_COMB.seq:*
5: /cgn1_7/ptodata/2/pna/US09_NEW_COMB.seq:*
6: /cgn1_7/ptodata/2/pna/US09_NEW_COMB.seq1:*
7: /cgn1_7/ptodata/2/pna/US09_NEW_COMB.seq2:*
8: /cgn1_7/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	594	100.0	3402	6	US-09-138-735-1	Sequence 1, Appli
	2	42	7.1	397803	6	US-09-803-110-217	Sequence 217, App
c	3	40.4	6.8	477	7	US-09-684-843A-1	Sequence 1, Appli
	4	38.8	6.5	657	7	US-09-922-480-3	Sequence 3, Appli
	5	38.8	6.5	657	7	US-09-923-236-3	Sequence 3, Appli
	6	38.8	6.5	657	7	US-09-922-469-3	Sequence 3, Appli
	7	38.6	6.5	411	7	US-09-902-540-5862	Sequence 5862, Ap
	8	38.6	6.5	1209	7	US-09-902-540-147	Sequence 147, App
	9	38.2	6.4	495	8	US-60-253-651-24827	Sequence 24827, A
	10	38.2	6.4	670	8	US-60-253-651-24775	Sequence 24775, A
	11	37.8	6.4	405	8	US-60-253-654-1232	Sequence 1232, Ap
	12	37.8	6.4	405	8	US-60-255-592-1232	Sequence 1232, Ap
	13	37.8	6.4	578	8	US-60-253-654-18403	Sequence 18403, A
	14	37.8	6.4	578	8	US-60-255-592-18403	Sequence 18403, A
	15	37.8	6.4	608	8	US-60-253-654-18342	Sequence 18342, A
	16	37.8	6.4	608	8	US-60-255-592-18342	Sequence 18342, A
	17	37.4	6.3	1039	8	US-60-278-561-9647	Sequence 9647, Ap
	18	37.4	6.3	6712	8	US-60-278-232-2818	Sequence 2818, Ap
c	19	37.2	6.3	522	6	US-09-823-241-9418	Sequence 9418, Ap
c	20	37.2	6.3	2406	6	US-09-760-466-345	Sequence 345, App
c	21	37.2	6.3	10733	7	US-09-902-540-1021	Sequence 1021, Ap
	22	36.8	6.2	435	5	US-09-842-827-1975	Sequence 1975, Ap
	23	36.8	6.2	435	7	US-09-912-292-40706	Sequence 40706, A
	24	36.8	6.2	621	7	US-09-840-424-56	Sequence 56, Appl
	25	36.6	6.2	30783	7	US-09-902-540-1258	Sequence 1258, Ap
c	26	36.4	6.1	1020	1	PCT-US01-08631-12729	Sequence 12729, A
	27	36.2	6.1	389	6	US-09-652-125A-2957	Sequence 2957, Ap
c	28	36.2	6.1	692	6	US-09-866-555-15847	Sequence 15847, A
c	29	36.2	6.1	2405	1	PCT-US01-08631-9988	Sequence 9988, Ap
	30	36	6.1	334	7	US-09-850-147-15625	Sequence 15625, A
	31	36	6.1	483	6	US-09-795-300-42	Sequence 42, Appl
	32	36	6.1	488	7	US-09-834-366-26649	Sequence 26649, A
	33	36	6.1	505	5	US-09-846-039-59	Sequence 59, Appl
	34	36	6.1	592	8	US-60-253-654-24123	Sequence 24123, A
	35	36	6.1	592	8	US-60-255-592-24123	Sequence 24123, A
	36	36	6.1	1448	5	US-09-925-298-318	Sequence 318, App
	37	36	6.1	16080	7	US-09-724-568-48	Sequence 48, Appl
	38	36	6.1	16080	7	US-09-724-569-48	Sequence 48, Appl
	39	36	6.1	16080	7	US-09-724-571-48	Sequence 48, Appl
	40	35.8	6.0	446	7	US-09-912-292-19829	Sequence 19829, A
	41	35.6	6.0	299	7	US-09-859-490-2340	Sequence 2340, Ap
	42	35.6	6.0	525	6	US-09-845-674-1322	Sequence 1322, Ap
	43	35.6	6.0	742	8	US-60-278-232-2812	Sequence 2812, Ap
	44	35.6	6.0	1089	6	US-09-235-810-75	Sequence 75, Appl
	45	35.6	6.0	1089	6	US-09-373-658-75	Sequence 75, Appl

ALIGNMENTS

RESULT 1
US-09-138-735-1
; Sequence 1, Application US/09138735
; GENERAL INFORMATION:
; APPLICANT: PARANHOS-BACCALA, Glaucia
; APPLICANT: LESENECHAL, Mylene
; APPLICANT: JOLIVET, Michel
; TITLE OF INVENTION: TRYPANOSOMA CRUZI ANTIGEN, GENE ENCODING THEREFOR AND METHODS
; TITLE OF INVENTION: DETECTING AND TREATING CHAGAS DISEASE
; FILE REFERENCE: WPB 36400B
; CURRENT APPLICATION NUMBER: US/09/138,735
; CURRENT FILING DATE: 1998-08-24
; PRIOR APPLICATION NUMBER: US 08/480,917
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: FR 94/10132
; PRIOR FILING DATE: 1994-08-12
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 3402
; TYPE: DNA
; ORGANISM: Trypanosoma cruzi
US-09-138-735-1

Query Match 100.0%; Score 594; DB 6; Length 3402;
Best Local Similarity 100.0%; Pred. No. 3.2e-131;
Matches 594; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 caggtacagcgtaacggcttttgcttcaatcgtacagccgacggtagctgcgtcctggct 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1232 caggtacagcgtaacggcttttgcttcaatcgtacagccgacggtagctgcgtcctggct 1291

Qy      61 gacatgtcgattcgattgacgatcttccatctccggtcctcccgcagggaagaacagcag 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1292 gacatgtcgattcgattgacgatcttccatctccggtcctcccgcagggaagaacagcag 1351

Qy     121 ccaggccaaaaaacatcggtagtggcgacggcgaaaccggggtgtgtgtcctcgggcact 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1352 ccaggccaaaaaacatcggtagtggcgacggcgaaaccggggtgtgtgtcctcgggcact 1411

Qy     181 gacgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatcccctgcatca 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1412 gacgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatcccctgcatca 1471

Qy     241 ccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggctgag 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1472 ccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggctgag 1531
```

```
Qy     301 ccgcacgtggggagcaagatcattgctaatctagtgaatcagctggggattaatgtcacc 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1532 ccgcacgtggggagcaagatcattgctaatctagtgaatcagctggggattaatgtcacc 1591

Qy     361 caaaggagcgtcgtcagcactggagcgccggccacgacgaggtctacggcggtgacgtcc 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1592 caaaggagcgtcgtcagcactggagcgccggccacgacgaggtctacggcggtgacgtcc 1651

Qy     421 acgactaccgccccgcagcgaacaagtccatacgggcacaatggccgacctgtgacggct 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1652 acgactaccgccccgcagcgaacaagtccatacgggcacaatggccgacctgtgacggct 1711

Qy     481 ggattggtggcagctaatagtggtgccagcgcggcctcgtctcccacagccgcggcgaaa 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1712 ggattggtggcagctaatagtggtgccagcgcggcctcgtctcccacagccgcggcgaaa 1771

Qy     541 ccaacaggagaagaaaaggcctccgcggcatgtgaaacgagctccgtggcgata 594
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1772 ccaacaggagaagaaaaggcctccgcggcatgtgaaacgagctccgtggcgata 1825
```

RESULT 2
US-09-803-110-217
; Sequence 217, Application US/09803110
; GENERAL INFORMATION:
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15490)D
; CURRENT APPLICATION NUMBER: US/09/803,110
; CURRENT FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/739,449
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 09/514,000
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: US 60/168,139
; PRIOR FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 13351
; SEQ ID NO 217
; LENGTH: 397803
; TYPE: DNA
; ORGANISM: Agrobacterium tumefaciens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(397803)
; OTHER INFORMATION: unsure at all n locations
US-09-803-110-217

Query Match 7.1%; Score 42; DB 6; Length 397803;
Best Local Similarity 60.5%; Pred. No. 1.8;
Matches 69; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

```
Qy     183 cgcggcgagtagcagtcataccaatacgacttctgccgctgctgcatccccctgcatcacc 242
           | | || | || |  ||  |     ||  || | || || || ||| ||||| || ||||
Db   66682 ccccgccaatatcgttcccaaactgacacctcccgcggcggcggcagcccctccagcacc 66741

Qy     243 ccctgtttcagcgccagccaaggcagccgcgcctcctgccgcggcgcgatcggc 296
               ||| | || || ||   || |   || ||||| |||||||| | | | ||
Db   66742 ggaagttgctgcaccggcagcggtacttgctcctccagccgcggctccacccgc 66795
```

RESULT 3
US-09-684-843A-1/c
; Sequence 1, Application US/09684843A
; GENERAL INFORMATION:
; APPLICANT: Ranum et al.
; TITLE OF INVENTION: SCA7 GENE AND METHODS OF USE
; FILE REFERENCE: Regents of the University of Minnesota
; CURRENT APPLICATION NUMBER: US/09/684,843A
; CURRENT FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/056,170
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 09/135,994
; PRIOR FILING DATE: 1998-08-18
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 477
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-684-843A-1

Query Match 6.8%; Score 40.4; DB 7; Length 477;
Best Local Similarity 53.9%; Pred. No. 1.9;
Matches 83; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

```
Qy     145 gcgacggcgaaaccggggtgtgtgtcctcgggcactgacgcggcgagtagcagtcatacc 204
           | | |||||    ||||   ||   || | ||  |||  | ||||        |  | |
Db     334 GTGGCGGCGGTGGCGGGTGCTGCTGCCGCTGGGGCTGCGGAGGCGGCGGCTGCTGCTGCT 275

Qy     205 aatacgacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaag 264
           | |  || |||| ||||||||  |  ||||  |  |  ||| | | ||  | ||
Db     274 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 215

Qy     265 gcagccgcgcctcctgccgcggcgcgatcggctg 298
           || || ||  || |||| || ||     | ||||
Db     214 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 181
```

RESULT 4
US-09-922-480-3
; Sequence 3, Application US/09922480
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Adler, David A.
; TITLE OF INVENTION: SECRETED SALIVARY ZSIG63 POLYPEPTIDE
; FILE REFERENCE: 97-71
; CURRENT APPLICATION NUMBER: US/09/922,480
; CURRENT FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/124,820
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 657
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Degenerate polynucleotide sequence for zsig63
; NAME/KEY: misc_feature
; LOCATION: (1)...(657)
; OTHER INFORMATION: n = A,T,C or G
US-09-922-480-3

Query Match 6.5%; Score 38.8; DB 7; Length 657;
Best Local Similarity 31.4%; Pred. No. 4.7;
Matches 89; Conservative 17; Mismatches 177; Indels 0; Gaps 0;

```
Qy     157 ccggggtgtgtgtcctcgggcactgacgcggcgagtagcagtcataccaatacgacttct 216
           ||  :  |  | | ::: |   | |  ||  |          |  | :    |     |
Db     340 ccnwsnmgnttyttywsngcngcngcngcnccngcngcnccnccnathgcngcngarccn 399

Qy     217 gccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaaggcagccgcgcct 276
           || || || ||  | :   |  |  | || ||  | ||   : |    || |  || ||
Db     400 gcngcngcngcnccnytnacngcnacnccngtngcngcngarccngcngcnggngcnccn 459

Qy     277 cctgccgcggcgcgatcggctgagccgcacgtggggagcaagatcattgctaatctagtg 336
             || || | :|   | || ||: | |  || ||     |:      ||  |
Db     460 gtngcngcngarccngcngcngargcnccngtnggngcngarccngcngcngargcnccn 519

Qy     337 aatcagctggggattaatgtcacccaaaggagcgtcgtcagcactggagcgccggccacg 396
                  | :      |              |  ||     | |  ||   :|   |
```

```
Db    520 gtngcngcngarccngcngcngargcnccngtnggngtngarccngcngcngargarccn 579

Qy    397 acgaggtctacggcggtgacgtccacgactaccgccccgcagc 439
          ::      |   : | |  ||  | | : |  | || ||  |:|
Db    580 wsnccngcngarccngcnacngcnaarccngcngcnccngarc 622


RESULT   5
US-09-923-236-3
; Sequence 3, Application US/09923236
; GENERAL INFORMATION:
;  APPLICANT: Sheppard, Paul O.
;  APPLICANT:  Adler, David A.
;  TITLE OF INVENTION: SECRETED SALIVARY ZSIG63 POLYPEPTIDE
;  FILE REFERENCE: 97-71
;  CURRENT APPLICATION NUMBER: US/09/923,236
;  CURRENT FILING DATE:  2001-08-03
;  PRIOR APPLICATION NUMBER: US 60/124,820
;  PRIOR FILING DATE: 1999-03-17
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
;   LENGTH: 657
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Degenerate polynucleotide sequence for zsig63
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(657)
;   OTHER INFORMATION: n = A,T,C or G
US-09-923-236-3


  Query Match             6.5%;  Score 38.8;  DB 7;  Length 657;
  Best Local Similarity   31.4%;  Pred. No. 4.7;
  Matches   89;  Conservative   17;  Mismatches  177;  Indels    0;  Gaps    0;

Qy    157 ccggggtgtgtgtcctcgggcactgacgcggcgagtagcagtcataccaatacgacttct 216
          ||  :  |  | | ::: |   | |  ||  |          |  | :    |     |
Db    340 ccnwsnmgnttyttywsngcngcngcngcnccngcngcnccnccnathgcngcngarccn 399

Qy    217 gccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaaggcagccgcgcct 276
          || || || ||  | :   |  |  | || ||  | ||   : |    || |  || ||
Db    400 gcngcngcngcnccnytnacngcnacnccngtngcngcngarccngcngcnggngcnccn 459

Qy    277 cctgccgcggcgcgatcggctgagccgcacgtggggagcaagatcattgctaatctagtg 336
             || || | :|   | || ||: | |  || ||      |:      ||  |
Db    460 gtngcngcngarccngcngcngargcnccngtnggngcngarccngcngcngargcnccn 519

Qy    337 aatcagctggggattaatgtcacccaaaggagcgtcgtcagcactggagcgccggccacg 396
                | :      |              | ||     | |  ||   :|   |
Db    520 gtngcngcngarccngcngcngargcnccngtnggngtngarccngcngcngargarccn 579

Qy    397 acgaggtctacggcggtgacgtccacgactaccgccccgcagc 439
          ::      |   : | |  ||  | | : |  | || ||  |:|
Db    580 wsnccngcngarccngcnacngcnaarccngcngcnccngarc 622


RESULT   6
US-09-922-469-3
; Sequence 3, Application US/09922469
; GENERAL INFORMATION:
;  APPLICANT: Sheppard, Paul O.
;  APPLICANT:  Adler, David A.
;  TITLE OF INVENTION: SECRETED SALIVARY ZSIG63 POLYPEPTIDE
;  FILE REFERENCE: 97-71
;  CURRENT APPLICATION NUMBER: US/09/922,469
;  CURRENT FILING DATE:  2001-08-03
;  PRIOR APPLICATION NUMBER: US 60/124,820
;  PRIOR FILING DATE: 1999-03-17
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
;   LENGTH: 657
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Degenerate polynucleotide sequence for zsig63
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(657)
;   OTHER INFORMATION: n = A,T,C or G
US-09-922-469-3


  Query Match             6.5%;  Score 38.8;  DB 7;  Length 657;
  Best Local Similarity   31.4%;  Pred. No. 4.7;
  Matches   89;  Conservative   17;  Mismatches  177;  Indels    0;  Gaps    0;

Qy    157 ccggggtgtgtgtcctcgggcactgacgcggcgagtagcagtcataccaatacgacttct 216
          ||  :  |  | | ::: |   | |  ||  |          |  | :    |     |
Db    340 ccnwsnmgnttyttywsngcngcngcngcnccngcngcnccnccnathgcngcngarccn 399

Qy    217 gccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaaggcagccgcgcct 276
          || || || ||  | :   |  |  | || ||  | ||   : |    || |  || ||
Db    400 gcngcngcngcnccnytnacngcnacnccngtngcngcngarccngcngcnggngcnccn 459

Qy    277 cctgccgcggcgcgatcggctgagccgcacgtggggagcaagatcattgctaatctagtg 336
             || || | :|   | || ||: | |  || ||      |:      ||  |
Db    460 gtngcngcngarccngcngcngargcnccngtnggngcngarccngcngcngargcnccn 519

Qy    337 aatcagctggggattaatgtcacccaaaggagcgtcgtcagcactggagcgccggccacg 396
                | :      |              | ||     | |  ||   :|   |
Db    520 gtngcngcngarccngcngcngargcnccngtnggngtngarccngcngcngargarccn 579

Qy    397 acgaggtctacggcggtgacgtccacgactaccgccccgcagc 439
          ::      |   : | |  ||  | | : |  | || ||  |:|
Db    580 wsnccngcngarccngcnacngcnaarccngcngcnccngarc 622


RESULT   7
US-09-902-540-5862
; Sequence 5862, Application US/09902540
; GENERAL INFORMATION:
;  APPLICANT: Goldman, Barry S.
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Wiegand, Roger C.
;  TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
;  FILE REFERENCE: 38-10(15849)B
;  CURRENT APPLICATION NUMBER: US/09/902,540
;  CURRENT FILING DATE:  2001-07-10
;  PRIOR APPLICATION NUMBER: 60/217,883
;  PRIOR FILING DATE: 2000-07-10
;  NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 5862
;   LENGTH: 411
;   TYPE: DNA
;   ORGANISM: Myxococcus xanthus
US-09-902-540-5862


  Query Match             6.5%;  Score 38.6;  DB 7;  Length 411;
  Best Local Similarity   52.7%;  Pred. No. 4.9;
  Matches  107;  Conservative    0;  Mismatches   94;  Indels    2;  Gaps    1;

Qy    215 ctgccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaaggcagccgcgc 274
          ||||  | ||||||  | | | | | |||||  | | |   |||  ||    | || |  | ||
Db     58 ctgctccggctgcaccgcattcgtcacctgccggtggtgcggcaagagaagctggtgggc 117

Qy    275 ctcctgccgcggcgcgatcggctgagccgcacgtggggagcaagatcattgctaatctag 334
          |||.|  | |  |||||  | |||| ||    ||  |  | | |   ||  |        |
Db    118 ctcatcacccaccgcgacctgctgcgc--gccgccgcaacccacgccacggacccggccg 175

Qy    335 tgaatcagctggggattaatgtcacccaaaggagcgtcgtcagcactggagcgccggcca 394
            | | |||  ||      | ||  |   | | ||| |||    || |    |||| ||
Db    176 cgcagccgctctgggccgcggacatcatgacgcgcgacgtgcagacggtgcgcccggaca 235
```

```
Qy    395 cgacgaggtctacggcggtgacg 417
          || ||  |     ||||||||||
Db    236 cgccgctgcgccgggcggtgacg 258
```

RESULT 8
US-09-902-540-147
; Sequence 147, Application US/09902540
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 147
; LENGTH: 1209
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-147

Query Match 6.5%; Score 38.6; DB 7; Length 1209;
Best Local Similarity 52.7%; Pred. No. 5.6;
Matches 107; Conservative 0; Mismatches 94; Indels 2; Gaps 1;

```
Qy    215 ctgccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaaggcagccgcgc 274
          ||||  | |||||| | | | | |||||  | | |   ||| ||    | || |  | ||
Db    285 ctgctccggctgcaccgcattcgtcacctgccggtggtgcggcaagagaagctggtgggc 344

Qy    275 ctcctgccgcggcgcgatcggctgagccgcacgtggggagcaagatcattgctaatctag 334
          ||| |  | |  ||||| | |||| ||    ||  |  | | |   ||  |       |
Db    345 ctcatcacccaccgcgacctgctgcgc--gccgccgcaacccacgccacggacccggccg 402

Qy    335 tgaatcagctggggattaatgtcacccaaaggagcgtcgtcagcactggagcgccggcca 394
          | | | |||  ||       | || |   | | ||| |||    || |     |||| ||
Db    403 cgcagccgctctgggccgcggacatcatgacgcgcgacgtgcagacggtgcgcccggaca 462

Qy    395 cgacgaggtctacggcggtgacg 417
          || ||  |     ||||||||||
Db    463 cgccgctgcgccgggcggtgacg 485
```

RESULT 9
US-60-253-651-24827
; Sequence 24827, Application US/60253651
; GENERAL INFORMATION:
; APPLICANT: Glenn, Mattew
; APPLICANT: Grigor, Murray R
; TITLE OF INVENTION: Compositions isolated from bovine tissue
; TITLE OF INVENTION: and methods for their use.
; FILE REFERENCE: 1055P3
; CURRENT APPLICATION NUMBER: US/60/253,651
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 27858
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 24827
; LENGTH: 495
; TYPE: DNA
; ORGANISM: Bovine
US-60-253-651-24827

Query Match 6.4%; Score 38.2; DB 8; Length 495;
Best Local Similarity 49.7%; Pred. No. 6.3;
Matches 97; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

```
Qy    106 agggaagaacagcagccaggccaaaaaaacatcggtagtggcgacggcgaaaccggggtgt 165
          |||||||  | |  | | |  |  |   |  |||  | ||| | |||| ||     |
Db    128 agggaagttccggcggcggctcctacggcggcgggcgcggctatggcggaagttccggcg 187

Qy    166 gtgtcctcgggcactgacgcggcgagtagcagtcataccaatacgacttctgccgctgct 225
          | | |  |||   ||  ||||||                |  |||| |  | |  ||| |
Db    188 gcggcggcggcggctacagcggcggaagctccagcggcggctacggcggcggaagctcca 247

Qy    226 gcatcccctgcatcacccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcg 285
          ||     | ||  |  |    ||| | ||| |||| | ||| || |   |||| |  |||
Db    248 gcggtggcggccacggcggtagttccggcggcagctacggcggcggaagctccagtggcg 307

Qy    286 gcgcgatcggctgag 300
          | |    |||| | |
Db    308 gtggccacggcggcg 322
```

RESULT 10
US-60-253-651-24775
; Sequence 24775, Application US/60253651
; GENERAL INFORMATION:
; APPLICANT: Glenn, Mattew
; APPLICANT: Grigor, Murray R
; TITLE OF INVENTION: Compositions isolated from bovine tissue
; TITLE OF INVENTION: and methods for their use.
; FILE REFERENCE: 1055P3
; CURRENT APPLICATION NUMBER: US/60/253,651
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 27858
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 24775
; LENGTH: 670
; TYPE: DNA
; ORGANISM: Bovine
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(670)
; OTHER INFORMATION: n = A,T,C or G
US-60-253-651-24775

Query Match 6.4%; Score 38.2; DB 8; Length 670;
Best Local Similarity 49.7%; Pred. No. 6.5;
Matches 97; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

```
Qy    106 agggaagaacagcagccaggccaaaaaaacatcggtagtggcgacggcgaaaccggggtgt 165
          |||||||  | |  | | |  |  |   |  |||  | ||| | |||| ||     |
Db    332 agggaagttccggcggcggctcctacggcggcgggcgcggctatggcggaagttccggcg 391

Qy    166 gtgtcctcgggcactgacgcggcgagtagcagtcataccaatacgacttctgccgctgct 225
          | | |  |||   ||  ||||||                |  |||| |  | |  ||| |
Db    392 gcggcggcggcggctacggcggcggaagctccagcggcggctacggcggcggaagctcca 451

Qy    226 gcatcccctgcatcacccccctgtttcagcgccagccaaggcagccgcgcctcctgccgcg 285
          ||     | ||  |  |    ||| | ||| |||| | ||| || |   |||| |  |||
Db    452 gcggtggcggccacggcggtagttccggcggcagctacggcggcggaagctccagtggcg 511

Qy    286 gcgcgatcggctgag 300
          | |    |||| | |
Db    512 gtggccacggcggcg 526
```

RESULT 11
US-60-253-654-1232
; Sequence 1232, Application US/60253654
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Norriss, Michael G.
; TITLE OF INVENTION: Compositions isolated from forage
; TITLE OF INVENTION: grasses and methods for their use.
; FILE REFERENCE: 1056P
; CURRENT APPLICATION NUMBER: US/60/253,654
; CURRENT FILING DATE: 2000-11-28

```
;  NUMBER OF SEQ ID NOS: 32581
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1232
;   LENGTH: 405
;   TYPE: DNA
;   ORGANISM: Lolium perenne
US-60-253-654-1232


  Query Match             6.4%;  Score 37.8;  DB 8;  Length 405;
  Best Local Similarity   46.2%;  Pred. No. 7.6;
  Matches  126;  Conservative    0;  Mismatches  147;  Indels    0;  Gaps    0;

Qy    153 gaaaccggggtgtgtgtcctcgggcactgacgcggcgagtagcagtcataccaatacgac 212
          | |  ||| |    || || ||   || | |||| |||  |  |   |  |||   |
Db     83 ggatacggagcacctgcccccgcctacggccgcgacgacgactacggacgccacgcccca 142

Qy    213 ttctgccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaaggcagccgc 272
            |  |  || | ||  ||||  || | ||  | ||  |  || |  | | ||  ||| |
Db    143 gcccccgcctacggcgcccccgccaccgcctacggtcgcgacgacgactacggacgccac 202

Qy    273 gcctcctgccgcggcgcgatcggctgagccgcacgtggggagcaagatcattgctaatct 332
          ||| ||  ||  | |     |||| |   ||     ||  |       ||        ||
Db    203 gcccccgccccggcctacggcggcggtcgcgacgacggctacggtcgccacgagccggct 262

Qy    333 agtgaatcagctggggattaatgtcacccaaaggagcgtcgtcagcactggagcgccggc 392
            |  |  | || |       |  ||  |       | |       |   ||| ||||
Db    263 ccggcctacgggggaggcggcggctacggagccccttgcccagcagcttacggcggcggc 322

Qy    393 cacgacgaggtctacggcggtgacgtccacgac 425
          | |||||| | ||||||| |  ||| ||  | |
Db    323 cgcgacgacggctacggccgccacgccccggcc 355


RESULT  12
US-60-255-592-1232
; Sequence 1232, Application US/60255592
; GENERAL INFORMATION:
;  APPLICANT: Glenn, Matthew
;  APPLICANT:  Norriss, Michael G.
;  TITLE OF INVENTION: Compositions isolated from forage
;  TITLE OF INVENTION:  grasses and methods for their use.
;  FILE REFERENCE: 1056P2
;  CURRENT APPLICATION NUMBER: US/60/255,592
;  CURRENT FILING DATE:  2000-12-14
;  NUMBER OF SEQ ID NOS: 32581
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1232
;   LENGTH: 405
;   TYPE: DNA
;   ORGANISM: Lolium perenne
US-60-255-592-1232


  Query Match             6.4%;  Score 37.8;  DB 8;  Length 405;
  Best Local Similarity   46.2%;  Pred. No. 7.6;
  Matches  126;  Conservative    0;  Mismatches  147;  Indels    0;  Gaps    0;

Qy    153 gaaaccggggtgtgtgtcctcgggcactgacgcggcgagtagcagtcataccaatacgac 212
          | |  ||| |    || || ||   || | |||| |||  |  |   |  |||   |
Db     83 ggatacggagcacctgcccccgcctacggccgcgacgacgactacggacgccacgcccca 142

Qy    213 ttctgccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaaggcagccgc 272
            |  |  || | ||  ||||  || | ||  | ||  |  || |  | | ||  ||| |
Db    143 gcccccgcctacggcgcccccgccaccgcctacggtcgcgacgacgactacggacgccac 202

Qy    273 gcctcctgccgcggcgcgatcggctgagccgcacgtggggagcaagatcattgctaatct 332
          ||| ||  ||  | |     |||| |   ||     ||  |       ||        ||
Db    203 gcccccgcccccggcctacggcggcggtcgcgacgacggctacggtcgccacgagccggct 262

Qy    333 agtgaatcagctggggattaatgtcacccaaaggagcgtcgtcagcactggagcgccggc 392
            |  |  | || |       |  ||  |       | |       |   ||| ||||
Db    263 ccggcctacgggggaggcggcggctacggagccccttgcccagcagcttacggcggcggc 322

Qy    393 cacgacgaggtctacggcggtgacgtccacgac 425
          | |||||| | ||||||| |  ||| ||  | |
Db    323 cgcgacgacggctacggccgccacgccccggcc 355


RESULT  13
US-60-253-654-18403
; Sequence 18403, Application US/60253654
; GENERAL INFORMATION:
;  APPLICANT: Glenn, Matthew
;  APPLICANT:  Norriss, Michael G.
;  TITLE OF INVENTION: Compositions isolated from forage
;  TITLE OF INVENTION:  grasses and methods for their use.
;  FILE REFERENCE: 1056P
;  CURRENT APPLICATION NUMBER: US/60/253,654
;  CURRENT FILING DATE:  2000-11-28
;  NUMBER OF SEQ ID NOS: 32581
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18403
;   LENGTH: 578
;   TYPE: DNA
;   ORGANISM: Festuca arundinaceae
US-60-253-654-18403


  Query Match             6.4%;  Score 37.8;  DB 8;  Length 578;
  Best Local Similarity   64.0%;  Pred. No. 7.9;
  Matches   57;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy    209 cgacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaaggcag 268
          ||||  |   |||| ||  | || ||  || | || || | |  |||| |||||   ||||
Db    330 cgacgtgctccgccgcctcgtcgccaacaccgccaccggctcgagcggcagccccgcag 389

Qy    269 ccgcgcctcctgccgcggcgcgatcggct 297
          || || |  | | || |||| |  |||||
Db    390 ccacgacggcggacggggcgagggcggct 418


RESULT  14
US-60-255-592-18403
; Sequence 18403, Application US/60255592
; GENERAL INFORMATION:
;  APPLICANT: Glenn, Matthew
;  APPLICANT:  Norriss, Michael G.
;  TITLE OF INVENTION: Compositions isolated from forage
;  TITLE OF INVENTION:  grasses and methods for their use.
;  FILE REFERENCE: 1056P2
;  CURRENT APPLICATION NUMBER: US/60/255,592
;  CURRENT FILING DATE:  2000-12-14
;  NUMBER OF SEQ ID NOS: 32581
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18403
;   LENGTH: 578
;   TYPE: DNA
;   ORGANISM: Festuca arundinaceae
US-60-255-592-18403


  Query Match             6.4%;  Score 37.8;  DB 8;  Length 578;
  Best Local Similarity   64.0%;  Pred. No. 7.9;
  Matches   57;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy    209 cgacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaaggcag 268
          ||||  |   |||| ||  | || ||  || | || || | |  |||| |||||   ||||
Db    330 cgacgtgctccgccgcctcgtcgccaacaccgccaccggctcgagcggcagccccgcag 389

Qy    269 ccgcgcctcctgccgcggcgcgatcggct 297
          || || |  | | || |||| |  |||||
Db    390 ccacgacggcggacggggcgagggcggct 418


RESULT  15
```

US-60-253-654-18342
; Sequence 18342, Application US/60253654
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Norriss, Michael G.
; TITLE OF INVENTION: Compositions isolated from forage
; TITLE OF INVENTION: grasses and methods for their use.
; FILE REFERENCE: 1056P
; CURRENT APPLICATION NUMBER: US/60/253,654
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 32581
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18342
; LENGTH: 608
; TYPE: DNA
; ORGANISM: Festuca arundinaceae
US-60-253-654-18342

Query Match 6.4%; Score 37.8; DB 8; Length 608;
Best Local Similarity 64.0%; Pred. No. 8;
Matches 57; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```
Qy    209 cgacttctgccgctgctgcatcccctgcatcacccccctgtttcagcgccagccaaggcag 268
          |||| |   |||| ||  | || ||  || | || || | |  |||| |||||   ||||
Db    330 cgacgtgctccgccgcctcgtcgccaacaccgccaccggctcgagcggcagcccccgcag 389

Qy    269 ccgcgcctcctgccgcggcgcgatcggct 297
          || || |  | | || |||| |  |||||
Db    390 ccacgacggcggacggggcgagggcggct 418
```

Search completed: September 21, 2001, 02:35:38
Job time: 14824 sec